Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 20 (2018), Issue 3 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Viewpoints

Original Papers

Short Paper

Letter to the Editor

Corrigenda and Addenda

Review

Information Architecture of Web-Based Interventions to Improve Health Outcomes: Systematic Review

Jillian Pugatch[1], MPH; Emily Grenen[1], MSc; Stacy Surla[1], MA; Mary Schwarz[2], BA; Heather Cole-Lewis[3], PhD, MPH

[1]ICF, Rockville, MD, United States

[2]ICF, Fairfax, VA, United States

[3]Johnson & Johnson Health and Wellness Solutions, Inc, New Brunswick, NJ, United States

Corresponding Author:
Jillian Pugatch, MPH
ICF
530 Gaither Road
Suite 500
Rockville, MD, 28770
United States
Phone: 1 301 572 0856
Email: jillian.pugatch@icf.com

Abstract

Background: The rise in usage of and access to new technologies in recent years has led to a growth in digital health behavior change interventions. As the shift to digital platforms continues to grow, it is increasingly important to consider how the field of information architecture (IA) can inform the development of digital health interventions. IA is the way in which digital content is organized and displayed, which strongly impacts users' ability to find and use content. While many information architecture best practices exist, there is a lack of empirical evidence on the role it plays in influencing behavior change and health outcomes.

Objective: Our aim was to conduct a systematic review synthesizing the existing literature on website information architecture and its effect on health outcomes, behavioral outcomes, and website engagement.

Methods: To identify all existing information architecture and health behavior literature, we searched articles published in English in the following databases (no date restrictions imposed): ACM Digital Library, CINAHL, Cochrane Library, Google Scholar, Ebsco, and PubMed. The search terms used included information terms (eg, information architecture, interaction design, persuasive design), behavior terms (eg, health behavior, behavioral intervention, ehealth), and health terms (eg, smoking, physical activity, diabetes). The search results were reviewed to determine if they met the inclusion and exclusion criteria created to identify empirical research that studied the effect of IA on health outcomes, behavioral outcomes, or website engagement. Articles that met inclusion criteria were assessed for study quality. Then, data from the articles were extracted using a priori categories established by 3 reviewers. However, the limited health outcome data gathered from the studies precluded a meta-analysis.

Results: The initial literature search yielded 685 results, which was narrowed down to three publications that examined the effect of information architecture on health outcomes, behavioral outcomes, or website engagement. One publication studied the isolated impact of information architecture on outcomes of interest (ie, website use and engagement; health-related knowledge, attitudes, and beliefs; and health behaviors), while the other two publications studied the impact of information architecture, website features (eg, interactivity, email prompts, and forums), and tailored content on these outcomes. The paper that investigated IA exclusively found that a tunnel IA improved site engagement and behavior knowledge, but it decreased users' perceived efficiency. The first study that did not isolate IA found that the enhanced site condition improved site usage but not the amount of content viewed. The second study that did not isolate IA found that a tailored site condition improved site usage, behavior knowledge, and some behavior outcomes.

Conclusions: No clear conclusion can be made about the relationship between IA and health outcomes, given limited evidence in the peer-reviewed literature connecting IA to behavioral outcomes and website engagement. Only one study reviewed solely manipulated IA, and we therefore recommend improving the scientific evidence base such that additional empirical studies investigate the impact of IA in isolation. Moreover, information from the gray literature and expert opinion might be identified and added to the evidence base, in order to lay the groundwork for hypothesis generation to improve empirical evidence on information architecture and health and behavior outcomes.

(J Med Internet Res 2018;20(3):e97) doi:10.2196/jmir.7867

KEYWORDS

information architecture; systematic review; health outcomes; behavior change; health behavior

Introduction

With the rise of new technology and digitization of our physical information environments, it is important to understand the role of digital information organization on user outcomes. This may be particularly important for the information architecture (IA) of Web content [1]. While no one definition of IA exists, it generally encompasses the organization of digital information, the labeling of information, and the navigation and search capabilities within a digital information space. The goal of IA is to build digital sites that enhance the user experience—in particular, the user's ability to find and use content [2].

IA is vital to website development. In commercial settings, good IA can enhance the ability of employees and customers to find information and decrease costs of Web redesign and maintenance [2]. However, IA is less often discussed in the context of digital spaces for behavior change and health outcomes. Moreover, IA best practices for commercial settings may not translate to health-related ones, where user needs are entirely different [2]. A user seeking information that is factual, concrete, and that they know exists (eg, the price of a new computer or the weekend forecast) will benefit from different site architecture than the user who wants to quit smoking or manage weight loss. In the latter scenarios, the information sought may be complicated and unfamiliar; the user may not even know exactly what information they should be seeking. Thus, while many IA recommendations exist, there is still a lack of empirical evidence for the role that IA plays in Web-based health behavior interventions.

Digital health interventions that mention IA primarily focus on navigation systems [3-5]. Generally, navigation systems concern the relationships among information or content at different levels—such as Web pages or sections. Structures can be hierarchical (top-down approach, with broader subjects encompassing smaller ones), matrix (movement along multiple dimensions), organic (free movement or exploration), or tunnel (sequential or linear organization) [6].

Many experts in the field recommend and implement a tunnel (or tunnel hybrid) design for behavior change websites. A recent systematic review of Web-based health intervention studies showed that tunneling structures were used in 90% of interventions reviewed. Of the interventions reviewed, all of those with a mental health focus used tunnel designs [7]. Users of websites with a tunnel design navigate in a sequential fashion to optimize the ordering of information and maximize the effectiveness of the site, in much the same way that one would read a novel or watch a television series from start to finish [1]. An example of tunnel design might be an online app that takes the user through a series of steps in a sequential order (eg, the app for health insurance on the American HealthCare.gov website), or a site with an e-learning module where lessons are presented in a predetermined order [1,8]. A tunnel experience is less likely to overwhelm users with information and options; it simplifies information consumption by defining what the user sees and when. In addition, tunnel design has the capacity to provide tailored "remedial" loops for users who do not pass certain knowledge test "check-points" or assessments [1]. In general, this type of feedback and reinforcement personalizes the experience and helps the individual progress through an intervention program. Evidence shows that personalized Web interventions are more efficacious in behavior change [1,9].

A hybrid design that includes elements of tunnel design provides an opportunity to give users more structure and guidance while also allowing a user to break free from a "locked" information structure if they so choose [1]. A website with a hybrid design might, for example, offer the user a table of contents that allows that individual to view website pages in any order. However, this same site might also include links within certain pages that direct users to a logical next step, thereby providing an element of tunnel design (eg, the National Institute of Justice's Laboratory Safety Training website) [10]. A hybrid tunnel design has the capacity to offer the user various ways of consuming the information, which may incentivize the user to take a more active role in their learning experience rather than simply turning pages, which is a risk with tunnel-only designs [1]. Hybrid design may also reduce attrition rate of a full tunnel design, as it does not deter individuals who may find the tunnel design too inflexible [1].

Conversely, free-form matrix—also known as organic—and hierarchical designs are less suitable for users unfamiliar with the content area (as is often the case for users of behavior change sites) because the freedom to explore information may make it difficult to navigate [1]. Additionally, these designs can make it more challenging for users to retrace their information search in order to review something previously seen [1].

Despite the aforementioned recommendations and the attention IA has received in the commercial sector, IA is largely a missed opportunity in the health behavior field. Most digital health intervention research describes the studies but fails to address the actual features of the Web tools being used, such as their IA [11-13]. Yet, understanding and implementing IA designs that best promote behavior change may be a simple and sustainable way to significantly improve the efficacy of digital interventions.

Thus, this review synthesizes the existing literature on website IA in the context of Web-based health interventions. We examine whether manipulating the information architecture of Web-based health interventions influences website use, health behaviors, and outcomes.

Methods

Inclusion and Exclusion Criteria

Articles were considered eligible if they met all inclusion criteria. In addition to being peer-reviewed and published in English, studies were included if they were (1) a randomized controlled trial (RCT), (2) an assessment of the effect of one type of IA compared to any other type of IA, (3) an intervention delivered in a Web-based setting, and (4) included either a primary health outcome measure (eg, disease status) or a secondary, proximal health outcome measure including change in knowledge, attitudes, or beliefs (eg, hepatitis knowledge) relating to the target health behavior, behavior change (eg, number of cigarettes smoked), website engagement (eg, number of pages visited), or attitudes towards the website (eg, perceived user control). No date restrictions were imposed. Interventions could address any health issue (eg, mental health, chronic conditions, and communicable diseases). Studies were included only if interventions were Web-based; interventions that focused on mobile apps or games, for example, were excluded.

Search Strategy

Literature searches were conducted on March 30, 2015. The following electronic databases were searched: ACM Digital Library, CINAHL, Cochrane Library, Google Scholar, Ebsco, and PubMed. The search terms used included information terms (eg, information architecture, interaction design, persuasive design), behavior terms (eg, health behavior, behavioral intervention, ehealth), and health terms (eg, smoking, physical activity, diabetes) (see Multimedia Appendix 1).

Eligibility assessment was performed independently by 2 reviewers. Disagreements between reviewers were resolved by consensus that included a third reviewer.

Data Extraction and Synthesis

A data extraction form was developed based on a priori categories established by 3 reviewers. Due to the small number of articles included in the review, this form was piloted on the three publications included in the systematic review.

Information was extracted from each included study on (1) characteristics of participants (including age, disease/behavior status), (2) type of intervention, (3) types of information architecture manipulated, (4) duration of the study, (5) website engagement outcomes, (6) knowledge, attitudes, and beliefs outcomes, and (7) health outcomes.

To determine the validity of eligible randomized trials, the pair of reviewers used the Cochrane Collaboration tool for assessing risk of bias in individual studies [14]. Disagreements in quality assessments were resolved by discussion between the 2 reviewers.

Results

Findings

Figure 1 illustrates the number of studies identified, screened, and included in this literature review [15]. The database literature search produced 782 citations. After duplicate citations were removed and the abstracts were reviewed, 17 citations met the inclusion criteria. The full text of these remaining citations were reviewed, and 14 were excluded because of study design (non-RCTs), a lack of IA manipulation, a lack of primary or secondary health outcome measures, or because they studied a non Web-based platform. Three articles were included in this systematic review [16-18].

Risk of Bias Assessment

The risk of bias criteria and outcomes are described in Table 1. Overall, risk of bias for all studies was low. Two studies failed to clearly report their method of random sequence generation [17,18], and one failed to report methods of allocation concealment and blinding of participants and personnel [17].

Study Characteristics

A summary of notable study characteristics is reported in Table 2. Sample sizes ranged from 561 [16] to 2523 [17] participants. One study was conducted in the United States [17], one in the Netherlands [18], and the third in Germany [16]. All studies were published in English.

Participants

Although all three studies lost participants to follow-up, only Weymann et al noted selective dropout, which occurred among participants with chronic lower back pain [16]. Those in the tailored condition were younger (mean 48.0, SD 12.9) and had higher education defined by having more than 10 years of education (119/190, 62.6%) than those in the control (age: mean 52.0, SD 12.7, P=.015; education: 94/188, 50.0%, P=.021). This study conducted intention-to-treat (ITT) and available cases (AC) analyses in order to determine the extent to which selection bias may have impacted the results.

Intervention Characteristics and Outcomes

The manipulation solely of IA was studied in only one publication [18], making it difficult to attribute the other two studies' results to the difference in IA. The two other studies manipulated website features (eg, interactivity, email prompts, and forums) and tailored content in addition to IA. All studies assessed some form of tunnel architecture against an organic architecture. Outcomes assessed included number of pages visited, time on site, website attrition, knowledge, perceived user control, perceived control, decisional conflict, patient empowerment, preparation for decision making, and change in knowledge.

Given that only one of the three publications assessed the isolated effect of IA, intervention characteristics and effect of IA on outcomes of interest are presented by study and categorized by whether the effect of IA was isolated. Table 3 includes more details regarding the studies' designs, results, and conclusions.

Interventions Assessing the Isolated Effect of Information Architecture

Crutzen et al Intervention Characteristics

The Crutzen et al study involved two versions of a website with different information architectures and a no-website control

group [18]. This publication assessed tunnel versus organic architecture.

One intervention group used a website about hepatitis with tunnel design. The pages on this site could be viewed only in a predetermined order and pages could not be skipped. The second group visited a freedom of choice (organic) site with identical content and the same number of pages as the tunnel version, but users had the ability to skip pages.

Crutzen et al Outcomes

Participants in the tunnel condition visited more pages (mean 11.4) compared to those in the freedom of choice condition (mean 7.4, $P<.001$). Users in the tunnel condition also spent more time on the site the than freedom of choice users (3:50 minutes compared to 2:38 minutes; $F_{1,452}=6.32$, $P=.01$).

Less user control had a negative effect on perceived website efficiency ($P<.01$), but a positive effect on knowledge gained ($P<.001$). Participants in the tunnel group scored higher on hepatitis knowledge compared with the freedom of choice group ($P<.001$).

Interventions Assessing the Non-Isolated Effect of Information Architecture

Danaher et al Intervention Characteristics

Danaher et al exposed smokeless tobacco users to a Basic and an Enhanced website for smokeless tobacco cessation called Chewfree.com [17]. The article assessed hybrid tunnel versus organic architecture. The Enhanced condition offered a tailored and interactive Web-based program that included text-based information, video-based testimonials, printable resources, interactive activities, annotated links to other website resources, and two Web forums. The Enhanced site used five navigational pages (one of which used a hybrid design that incorporated tunneling). The Basic condition represented a subset of the content presented in the Enhanced condition and included text-based content using four navigational pages. It also offered a printable self-help smokeless tobacco cessation booklet, printable cessation resource, and annotated links to other recommended websites for tobacco cessation.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram. RCT: randomized controlled trial; IA: information architecture.

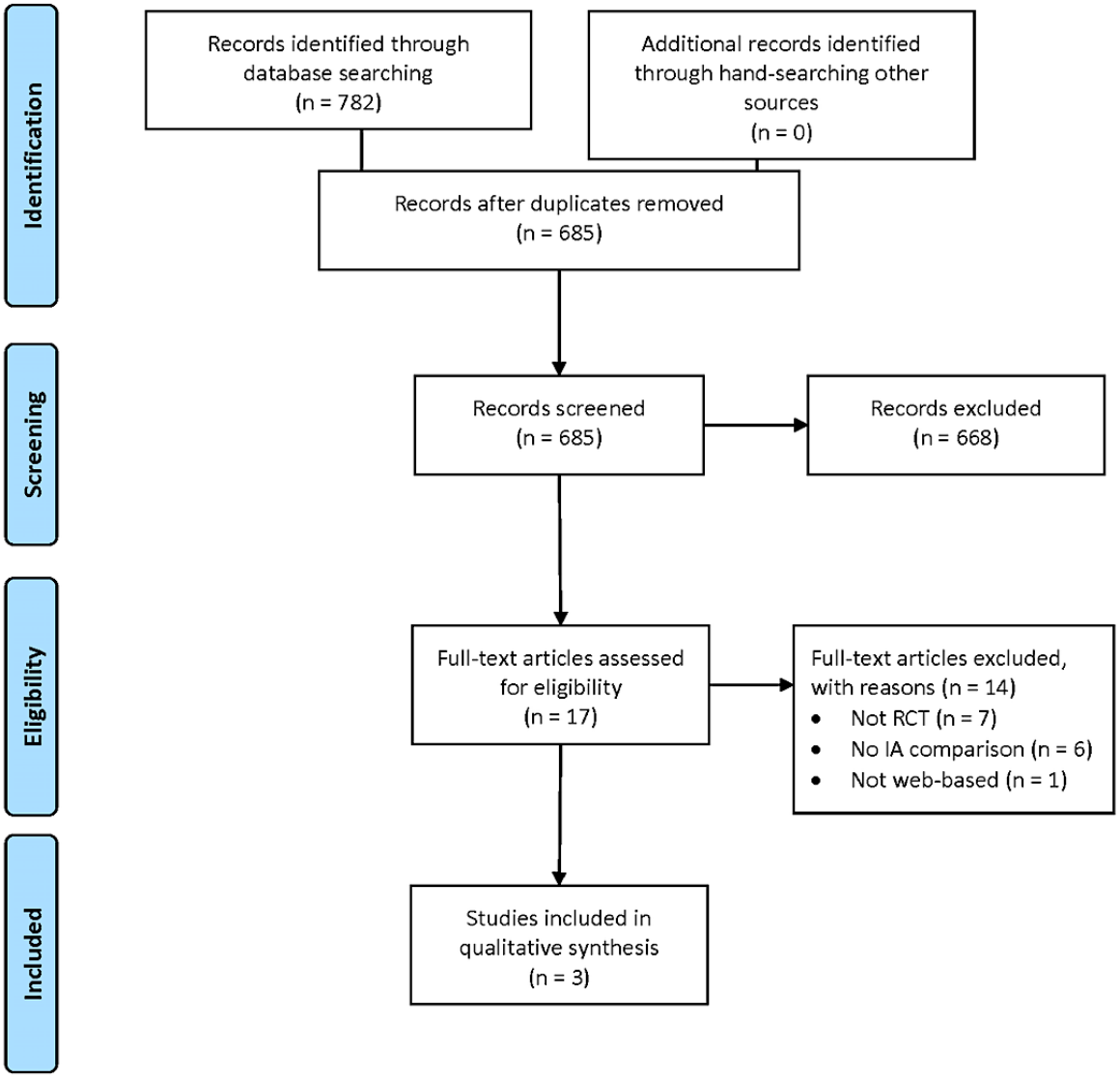


Table 1. Risk of bias assessment[a] for individual studies.

Criteria	Weymann et al [16]		Danaher et al [17]		Crutzen et al [18]	
	Risk of bias	Support for judgment	Risk of bias	Support for judgment	Risk of bias	Support for judgment
Random sequence generation	Low	Simple randomization of participants performed by a software program	Unclear	No description of the methodology to generate allocation sequence	Unclear	No description of the methodology to generate allocation sequence
Allocation concealment	Low	Randomization software assured the concealment of allocation.	Unclear	No description of methods used to generate intervention or control allocations. However, given the study took place online, it is unlikely that participants would be aware of allocation.	Low	No description of methods used to generate intervention or control allocations. However, participants were not informed about the existence of these 3 groups or that the study focus was on website use.
Blinding of participants and personnel	Low	Participants were aware that there were intervention and control groups, but blinded to their assignment. However, authors stated "it might be possible that participants identified the intervention group due to the unusual dialogue-based delivery format used in the intervention group." Due to software-automated allocation, personnel remained blinded.	Unclear	No description of participant or researcher blinding. However, given study took place online, it is unlikely that either participants or researchers would have been aware of the intervention and control allocations.	Low	No description of participant or researcher blinding. However, given the study took place online, it is unlikely that either participants or researchers would have been aware of intervention and control allocations.
Blinding of outcomes assessment	Low	Outcomes all collected via self-reported questionnaires.	Low	Website use and engagement comprised all outcomes measured, which were calculated via an automated computer program.	Low	All outcomes were collected via a computer server (website use) or via self-reported questionnaires completed online.
Incomplete outcome data	Medium	Although authors stated that "attrition was comparatively low for an online trial," they found evidence for selective dropout between the control and intervention conditions.	Low	Only about 6% of participants were not included in the analyses. This included participants who never visited their assigned website or returned only to complete online assessments.	Low	Relatively small dropout rate between pre-test and follow-up, and authors reported that there was neither selective dropout nor a difference in dropout between conditions.
Selective reporting	Low	Data for all outcomes described in the study protocols[b,c] were reported.	Low	Data from all outcomes indicated in the Methods section reported in the Results section.	Low	Data from all outcomes indicated in the Methods section reported in the Results section.
Other bias	Low	None identified.	Low	None identified.	Low	None identified.

[a]Risk of bias was categorized as low, medium, or high based on whether reviewers thought the methods or descriptions indicated a low, medium, or high risk. "Unclear" risk of biases was noted for studies that lacked a description of that domain.

[b][19].

[c][20].

Danaher et al Outcomes

Specific tunnel elements and IA were not isolated in this intervention. The Enhanced site generated more usage. Participants in the enhanced condition made more visits and spent more time on the site than participants in the Basic condition ($P<.001$). Users in the Enhanced condition continued to use the site for more days than Basic website users ($P<.001$). Interestingly, more cessation content was visited in the Basic condition, though the study authors note that this could be due to the same content being difficult to find in the Enhanced site.

Weymann et al Intervention Characteristics

Weymann et al compared a tailored and interactive site with some tunneling elements to a control site without tunneling [16]. The study assessed dialogue-based and tunnel versus organic architecture. There were intervention and control sites for people with type 2 diabetes and chronic lower back pain for a total of four conditions. The look of the websites (colors, font, figures, and pictures) was identical in all conditions, and participants could view sites as often as they wished.

In the tailored conditions, the delivery format was a dialogue-based, tunnel design. The dialogue aspect of the design attempts to imitate a conversation with a health professional. Various check-points assessed user knowledge and attitude toward a topic, and content was then modified according to their answer. Users were given limited control over the sequence in which they viewed content—although they were permitted to pick from one of several options at the end of each text passage. On the control websites, the content was not tailored and was not presented in a dialogue format. In contrast to the tailored, interactive version, the control website users were given freedom to view content in any order by selecting topics from a menu.

Weymann et al Outcomes

Analyses and findings of this study did not explore IA specifically. In this study, the tailored and tunnel conditions spent more time on the site (51.16 minutes) than the control groups (37.6 minutes) (P<.001). Results for the other outcomes are as follows:

- Knowledge after the first visit ITT: no significant difference (P=.53); AC: tailored group had significantly more knowledge (P=.02) than control
- Patient empowerment ITT: no significant difference; AC: tailored group had better emotional well-being (a subscale of empowerment) than control (P=.009)
- Decisional conflict disease main effect for ITT and AC
- Preparation for decision making ITT: no significant difference; AC: disease main effect (P=.02)

Content tailoring and interactivity may increase knowledge and reduce health-related negative effects in persons who use interactive health communication apps.

Table 2. Characteristics of included studies.

Study	Intervention arms	Population	Sample size	Health concern	Outcome measures[a]
Weymann et al [16]	Tunnel condition: Tunnel design and tailored content Control condition: Free-form navigation website with untailored content not presented in a dialogue format	Adults in Germany with access to internet and sufficient computer/internet literacy. Participants had either a self-reported diagnosis of type 2 diabetes or chronic low back pain.	Baseline (n=561): Tunnel condition n=283; Control condition n=278 Follow-up (n=295): Tunnel condition n=146; Control condition n=149	Type 2 diabetes; Chronic lower back pain	(1) Time on website, (2) Knowledge after first website visit, (3) Decisional conflict after 1st website visit, (4) Preparation for decision making after 1st website visit, (5) Patient empowerment at 3-month follow-up
Danaher et al [17]	Enhanced condition: Hybrid tunnel design website with interactive, tailored, rich media Control condition: Static, text-based website with free navigation to all content	Adult smokeless tobacco users in the United States.	Baseline (n=2523): Enhanced condition n=1260; Control condition n=1263 Follow-up (n=2375): Enhanced condition n=1200; Control condition n=1175	Smokeless tobacco use	(1) Website visits at T1, T2, and T3, (2) Time on website at T1, T2, and T3, (3) Website attrition from T1-T3
Crutzen et al [18]	Tunnel condition: Website with tunnel design and less user control Free-form condition: Freedom of choice design where users had ability to skip pages Control condition: No exposure to website	Adult internet users in the Netherlands.	Baseline (n=668): Tunnel condition: n=226; Free-form condition: n=228; Control: n=214 Follow-up (n=571): Tunnel condition: n=200; Free-form condition: n=193; Control: n=178	Hepatitis	(1) Time on website at T0, (2) Number of pages viewed at T0, (3) Perceived user control at T1, (4) User perceptions at T1, (4) Change in hepatitis knowledge from T0-T2

[a]T0=baseline, T1=time 1, T2=time 2, and T3=time 3, when user data were collected.

Table 3. Results and conclusions of included studies.

Author	Data collection points	Website use results	Knowledge, attitudes, beliefs results	Conclusion
Weymann et al [16]	T1: Immediately after 1st website visit, T2: 3-month follow-up	Time on website: Tunnel condition mean 51.2 min; Control condition mean 37.6 min (P<.001)	Knowledge after 1st visit: ITT[a] analysis=Tailored condition mean 77.9; Control condition mean 76.3 (P=.53). AC[b] analysis=Tailored condition mean 79.1; Control condition mean 75.2 (P=.02) Decisional conflict after 1st visit: No significant intervention main effects for AC or ITT analyses. Preparation for decision making after 1st visit: No significant intervention main effects for AC or ITT analyses. Patient empowerment at 3-month follow-up: ITT analysis=No significant intervention main effect or interaction. AC analysis=Intervention main effect for Emotional Well-being (subscale of patient empowerment). Tailored condition mean 68.5; Control condition mean 60.0 (P=.009).	Participants spent more time with tunnel site than the control. In the ITT analyses, this did not result in more knowledge or empowerment. Sensitivity analyses (AC) showed that participants in tunnel condition displayed more knowledge and emotional well-being. However, on other measures of patient empowerment, there was no difference between the 2 conditions.
Danaher et al [17]	T1: 6 weeks after enrollment, T2: 3 months after enrollment, T3: 6 months after enrollment	Website visits: Enhanced condition made more visits (z=-16.64, P<.001, 2-tailed). Time on website: Enhanced condition spent more time viewing website content (z=-17.63, P<.001, 2-tailed). Website attrition: Enhanced condition showed slower attrition (P<.001 for both log-rank and Breslow tests).	N/A[c]	Study suggests that hybrid tunnel IA may encourage higher participant engagement with website content than free-form IA. Engagement measures are important in understanding program effectiveness. However, the study is limited in that it does not directly measure behavioral outcomes.
Crutzen et al [18]	T0: Pretest, T1: Immediately after viewing website, T2: 1 week after viewing website	Time on website: Tunnel condition mean 3:50 min; Free-form condition mean 2:38 min ($F_{1,452}$=6.32, P=.01). Number of pages visited: Tunnel condition mean 11.4 pages; Free-form condition mean 7.4 pages ($F_{1,452}$=171.49, P<.001).	Perceived control: Free-form condition higher mean 5.2; Tunnel condition mean 3.9 ($F_{1,452}$=134.32, I<.001) Change in hepatitis knowledge: Tunnel condition pretest mean 5.0, posttest mean 8.2; Free-form pretest mean 5.4, posttest mean 7.2; Control condition pretest mean 5.4, posttest mean 5.6 ($F_{2,567}$=47.24, P<.001). All pairwise comparisons significant (P<.001).	IA that provides less choice may improve intervention engagement and disease knowledge, which may benefit health behavior outcomes. However, user perceptions of efficiency may be compromised by restricting user choice.

[a]ITT: intention-to-treat.

[b]AC: available cases.

[c]N/A: not applicable.

Discussion

Principal Findings

Given the limited body of evidence connecting IA to behavioral outcomes and website engagement, no clear conclusions can be made about the relationship between IA and health outcomes. Moreover, several weaknesses in the design of the studies identified make it challenging to generalize results. Only one of the articles, for example, explicitly and empirically manipulated IA by itself [18]. The other studies included other manipulations to website features and tailoring, making it difficult to attribute the results to the difference in IA. Loss to follow-up also makes it difficult to determine whether outcomes resulted from the intervention itself or simply bias [16]. Future empirical research on IA necessitates more robust study designs that isolate the effect of IA and minimize loss to follow-up. Adopting a more nuanced study design approach may even allow researchers to isolate IA while testing other features of an intervention in an RCT. For example, the Sequential Multiple Assignment Randomized Trials (SMART) design for adaptive interventions—in which participants move through multiple

stages of an intervention and get reassigned to several intervention options—might offer an opportunity to test IA features tailored to particular users depending on their behaviors and needs within the context of a larger intervention trial [21].

The publications in this review did not assess health outcomes—instead they focused on more proximal outcomes such as behavior change and website engagement. It is generally accepted that some level of engagement with a digital intervention is necessary in order to achieve any benefit [22,23], and as such, engagement is often used as a proxy indicator of behavior change or health outcomes. However, engagement measures are not as robust as behavior change or health outcomes (longer engagement might, for example, reflect difficulties in understanding or navigating through the site), and health practitioners and clinicians should collaborate with developers to conduct randomized trials with health outcomes, in order to improve the body of literature on IA.

Strengths and Limitations

There were some limitations to the review process itself. First, we did not prospectively register our literature review, thereby risking duplication. Also, because search terms attempted to capture websites relating to such a broad topic (ie, health behavior change), it is possible that despite the long list of health terms included, we missed relevant IA and heath behavior change studies.

Despite these limitations, the results of the Crutzen et al study do suggest that less user control (ie, tunnel design) may increase website use and knowledge gained [18]. Less user control may have more impact in a health behavior change context for a variety of reasons [1]. First, a tunnel experience may avoid overwhelming users with too many options by controlling what the user sees and when. In addition, tunnel design can provide a more tailored user experience by tracking users' progress and knowledge attainment (via tests or assessments) and delivering appropriate content accordingly [1]. Feedback and reinforcement not only personalize the user experience but also help the individual progress through an intervention program and adopt a behavior change [1,9]. The Weymann et al study included some of these tailoring elements; it is certainly possible that this attributed to the higher engagement levels in the intervention group.

Future Considerations

More research is needed to explore whether tunneling can improve user engagement and knowledge and to understand how it impacts behavior outcomes. Additionally, the studies identified here examine only navigation systems (specifically, tunneling versus organic design), which is just one component of information architecture. Future research should consider the effectiveness of other IA organizing designs (ie, hierarchical and matrix design), as well as other IA elements, such as labeling systems (ie, how information is represented). Some research is beginning to explore the effect of enhanced search systems (ie, how users look for information) within a health-related website [24].

Future reviews might also consider a larger scope of literature. For our purposes, we considered only peer-reviewed RCTs. However, there may be a body of gray literature, albeit less robust, on the subject of IA for Web-based health interventions that could be worth investigating given the lack of evidence found here.

The limited evidence base found in this review demonstrates that IA is a largely unstudied aspect of the health behavior field. If a robust evidence base is established and effective IA designs for health behavior change are identified, the development of Web-based interventions could be streamlined. In addition to improved intervention efficacy, evidence-based IAs could free up resources like time and money to enhance other aspects of the intervention such as graphic design, user experience, marketing, or evaluation. Also, the use of A/B or pre-post testing through automated digital platforms could make building an evidence base more feasible.

Conclusion

Due to the limited evidence base, few claims can be made about the relationship between IA and health and behavior outcomes. There is support for the effect of tunneling on user engagement and knowledge, but more research is needed to support this claim.

This synthesis of information will provide guidance to practitioners designing websites for health behavior and health outcomes. We hope this serves as a starting point for hypothesis generation to improve empirical evidence on IA and health and behavior outcomes.

Acknowledgments

This work was funded by the National Cancer Institute's Tobacco Control Research Branch of the National Institutes of Health, National Cancer Institute contract #HHSN261201400002B and #HHSN26100007. National Cancer Institute staff did not coauthor this manuscript but did have a limited role in study design and decision to publish.

The authors would like to thank Erik Augustson, PhD, MPH, and Yvonne Hunt, PhD, MPH, for their contributions to the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search terms.

[PDF File (Adobe PDF File), 71KB - jmir_v20i3e97_app1.pdf]

References

1. Danaher B, McKay H, Seeley J. The information architecture of behavior change websites. J Med Internet Res 2005 May 18;7(2):e12 [FREE Full text] [doi: 10.2196/jmir.7.2.e12] [Medline: 15914459]
2. Rosenfeld L, Morville P, Arango J. In: Rufino A, editor. Information Architecture for the Web and Beyond. 4th edition. Sebastopol, CA: O'Reilly Media, Inc; 2015:1-461.
3. Swartz LHG, Noell JW, Schroeder SW, Ary DV. A randomised control study of a fully automated internet based smoking cessation programme. Tob Control 2006 Feb;15(1):7-12 [FREE Full text] [doi: 10.1136/tc.2003.006189] [Medline: 16436397]
4. Boucher S, Edwards O, Gray A, Nada-Raja S, Lillis J, Tylka T, et al. Teaching Intuitive Eating and Acceptance and Commitment Therapy Skills Via a Web-Based Intervention: A Pilot Single-Arm Intervention Study. JMIR Res Protoc 2016 Oct 14;5(4):e180-e186 [FREE Full text] [doi: 10.2196/resprot.5861] [Medline: 27742602]
5. Danielson CK, McCauley JL, Gros KS, Jones AM, Barr SC, Borkman AL, et al. SiHLEWeb.com: Development and usability testing of an evidence-based HIV prevention website for female African-American adolescents. Health Informatics J 2016 Jun;22(2):194-208 [FREE Full text] [doi: 10.1177/1460458214544048] [Medline: 25167865]
6. Garrett J. In: Nolan MJ, editor. The Elements of User Experience: User-Centered Design for the Web and Beyond. 2nd edition. Berkeley, CA: New Riders; 2011:1-174.
7. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JEWC. 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]
8. National Cancer Institute. Quiz: Are You Depressed?. 2017. URL: https://smokefree.gov/quiz-are-you-depressed [accessed 2017-12-05] [WebCite Cache ID 6vU8z8hHY]
9. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. Ann Behav Med 2009 Aug;38(1):18-27 [FREE Full text] [doi: 10.1007/s12160-009-9133-4] [Medline: 19802647]
10. National Institute of Justice. 2017. Training: Laboratory Safety Programs URL: labsafety.training.nij.gov [accessed 2018-03-14] [WebCite Cache ID 6xuwOlqq5]
11. Greenhalgh T, Russell J. Why do evaluations of eHealth programs fail? An alternative set of guiding principles. PLoS Med 2010 Nov 02;7(11):e1000360 [FREE Full text] [doi: 10.1371/journal.pmed.1000360] [Medline: 21072245]
12. World Health Organization. New horizons for health through mobile technologies (Global Observatory for eHealth Series, Volume 3). Geneva, Switzerland: World Health Organization; 2011:1-102.
13. Adibi S, editor. Mobile Health: A technology roadmap. Switzerland: Springer; 2015:1-1172.
14. Higgins J. Cochrane Handbook for Systematic Reviews of Interventions Version 5. 2011. URL: http://handbook-5-1.cochrane.org/ [accessed 2017-12-05] [WebCite Cache ID 6vUC9bJwf]
15. Moher D, Liberati A, Tetzlaff J, Altman D, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-269, W64. [Medline: 19622511]
16. Weymann N, Dirmaier J, von Wolff A, Kriston L, Härter M. Effectiveness of a Web-based tailored interactive health communication application for patients with type 2 diabetes or chronic low back pain: randomized controlled trial. J Med Internet Res 2015 Mar 03;17(3):e53 [FREE Full text] [doi: 10.2196/jmir.3904] [Medline: 25736340]
17. Danaher BG, Boles SM, Akers L, Gordon JS, Severson HH. Defining participant exposure measures in Web-based health behavior change programs. J Med Internet Res 2006 Aug 30;8(3):e15 [FREE Full text] [doi: 10.2196/jmir.8.3.e15] [Medline: 16954125]
18. Crutzen R, Cyr D, de Vries NK. The role of user control in adherence to and knowledge gained from a website: randomized comparison between a tunneled version and a freedom-of-choice version. J Med Internet Res 2012 Mar 09;14(2):e45 [FREE Full text] [doi: 10.2196/jmir.1922] [Medline: 22532074]
19. Dirmaier J, Härter M, Weymann N. A tailored, dialogue-based health communication application for patients with chronic low back pain: study protocol of a randomised controlled trial. BMC Med Inform Decis Mak 2013 Jun 14;13:66 [FREE Full text] [doi: 10.1186/1472-6947-13-66] [Medline: 23768119]
20. Weymann N, Härter M, Dirmaier J. A tailored, interactive health communication application for patients with type 2 diabetes: study protocol of a randomised controlled trial. BMC Med Inform Decis Mak 2013 Feb 13;13:24 [FREE Full text] [doi: 10.1186/1472-6947-13-24] [Medline: 23406466]
21. Lei H, Nahum-Shani I, Lynch K, Oslin D, Murphy SA. A "SMART" design for building individualized treatment sequences. Annu Rev Clin Psychol 2012;8:21-48 [FREE Full text] [doi: 10.1146/annurev-clinpsy-032511-143152] [Medline: 22224838]
22. 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 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
23. Richardson A, Graham AL, Cobb N, Xiao H, Mushro A, Abrams D, et al. Engagement promotes abstinence in a web-based cessation intervention: cohort study. J Med Internet Res 2013 Jan 28;15(1):e14 [FREE Full text] [doi: 10.2196/jmir.2277] [Medline: 23353649]

24. Zeng Y, Liu X, Wang Y, Shen F, Liu S, Rastegar-Mojarad M, et al. Recommending Education Materials for Diabetic Questions Using Information Retrieval Approaches. J Med Internet Res 2017 Oct 16;19(10):e342 [FREE Full text] [doi: 10.2196/jmir.7754] [Medline: 29038097]

Abbreviations

AC: available cases
IA: information architecture
ITT: intention to treat
RCT: randomized controlled trial
T0, T1, T2, T3: baseline, time 1, time 2, time 3

Edited by G Eysenbach; submitted 19.04.17; peer-reviewed by A Riis, Y Wang, M Ashford, A Cyr; comments to author 12.07.17; revised version received 06.12.17; accepted 03.01.18; published 21.03.18.

Please cite as:
Pugatch J, Grenen E, Surla S, Schwarz M, Cole-Lewis H
Information Architecture of Web-Based Interventions to Improve Health Outcomes: Systematic Review
J Med Internet Res 2018;20(3):e97
URL: http://www.jmir.org/2018/3/e97/
doi:10.2196/jmir.7867
PMID:29563076

Viewpoint

Social Media as a Catalyst for Policy Action and Social Change for Health and Well-Being: Viewpoint

Douglas Yeung[1], PhD

RAND Corporation, Santa Monica, CA, United States

Corresponding Author:
Douglas Yeung, PhD
RAND Corporation
1776 Main Street
Santa Monica, CA,
United States
Phone: 1 3103930411
Email: dyeung@rand.org

Abstract

This viewpoint paper argues that policy interventions can benefit from the continued use of social media analytics, which can serve as an important complement to traditional social science data collection and analysis. Efforts to improve well-being should provide an opportunity to explore these areas more deeply, and encourage the efforts of those conducting national and local data collection on health to incorporate more of these emerging data sources. Social media remains a relatively untapped source of information to catalyze policy action and social change. However, the diversity of social media platforms and available analysis techniques provides multiple ways to offer insight for policy making and decision making. For instance, social media content can provide timely information about the impact of policy interventions. Social media location information can inform where to deploy resources or disseminate public messaging. Network analysis of social media connections can reveal underserved populations who may be disconnected from public services. Machine learning can help recognize important patterns for disease surveillance or to model population sentiment. To fully realize these potential policy uses, limitations to social media data will need to be overcome, including data reliability and validity, and potential privacy risks. Traditional data collection may not fully capture the upstream factors and systemic relationships that influence health and well-being. Policy actions and social change efforts, such as the Robert Wood Johnson Foundation's effort to advance a culture of health, which are intended to drive change in a network of upstream health drivers, will need to incorporate a broad range of behavioral information, such as health attitudes or physical activity levels. Applying innovative techniques to emerging data has the potential to extract insight from unstructured data or fuse disparate sources of data, such as linking health attitudes that are expressed to health behaviors or broader health and well-being outcomes.

(J Med Internet Res 2018;20(3):e94) doi:10.2196/jmir.8508

KEYWORDS

social media; health policy; health promotion; health knowledge, attitudes, practice; social change

Introduction

Recent efforts to improve health and well-being have looked beyond medicine and health care to consider the influence of a network of upstream factors such as social connectedness, civic engagement, and the physical environment. However, traditional data collection may not fully capture the upstream factors and systemic relationships that could drive positive change in health and health care in the United States [1]. Furthermore, measuring community attitudes and behaviors to inform policy is frequently accomplished with labor-intensive surveys or interviews. Policy actions that are intended to influence broader social change in upstream drivers, such as the Robert Wood Johnson Foundation's effort to advance a Culture of Health, will need to incorporate a more varied range of behavioral information to inform those actions, when timely or relevant information is otherwise unavailable. Applying innovative techniques to emerging data has the potential to extract insight from unstructured data or fuse disparate sources of data, such as linking health attitudes that are expressed to health behaviors or broader outcomes.

Social media provides an unprecedented opportunity to understand values and expectations about health, and to track healthy behaviors and outcomes in timely ways. Using Web-based applications, people now create and share a wide range of content that may provide richer insight into the value

they place on health and well-being for themselves, their friends and family, and the surrounding community. For instance, the amount of discussion (eg, frequency of Twitter mentions) about healthy eating, physical activities, or stress management suggests how much people are thinking about well-being or engaging in health promotion activities [2]. Social media check-ins, that is, when people post their location at a certain place (eg, at a public park or restaurant), can reveal how often people use healthy places.

These behaviors offer insights into how social media can reflect what people consider important, that is, the topics they consider worth discussing or activities they consider worth their participation. Tracking these health-related conversations or actions over time may provide early indicators about important health events or reveal activity patterns that contribute to, or detract from, health and well-being. Policy makers can then use this insight to inform either targeted interventions or broader, longer-term initiatives. Despite a wide range of academic research on how to categorize and mine social media information, social media remains relatively untapped as a source of information to catalyze policy action and advance social change, specifically for health and well-being. Harnessing this potential will require careful consideration to establish validity and reliability, such as addressing bias in either social media usage or analytic techniques.

Why Is Social Media Well Suited to Support Policy Action and Social Change for Health and Well-Being?

Rapidly refreshed and constantly changing, social media data can help track attitudes and behaviors along multiple paths toward improving health and well-being. Current policy making to advance cultural change in health and well-being has emphasized multisectoral collaborative efforts. For example, in the Culture of Health framework, there is a focus on a dynamic process of improving population well-being, reliant on large amounts of information that updates rapidly and differs by location, as well as by geographic, demographic, and social sectors. However, traditional public policy measurements, such as tracking health policies, community will, or other community or infrastructure level indicators fall short in providing the depth needed on behavioral insights.

Social media data can fit the bill. They are both communal (ie, containing shared ties and social connections) and individualistic (ie, highly granular) and can be broken down to examine behaviors across geography, demographics, or socioeconomic status. This is because although social media is widely used, it is also heavily used by certain segments of the population (eg, youth, minorities) that may be important for issues of health equity. In a Culture of Health, shifting and influencing mindset and expectations about health and well-being is central to social change [3]. Social media could be a particularly valuable tool to capture the mindset and expectations of people in these groups. For instance, minorities may be more likely than other groups to access the Internet and social media primarily through their mobile phones, and these actions could be better tracked to understand health and well-being behaviors [4].

Data from social media can provide insight into whether people exhibit similar views and behaviors around the importance of health and well-being. This is because social media, and big data more broadly, offer a unique type of naturalistic, behavioral data that are a rich source of information on health attitudes and behaviors. For instance, the amount of online discussion (ie, frequent word use) on a given topic may be related to general interest in that topic, particularly when considering text and sentiment analysis. One study of postings on a weight-loss blog suggested that sharing one's negative emotions, as indicated by the use of sadness words, was linked to greater success in losing weight [5]. Similar explorations of social media as a way to understand health attitudes and behavior (eg, [6-8]) and track health outcomes (eg, [9,10]) further illustrate the potential in exploring social media data to establish their utility for policy uses.

Social Media Offers Multiple Pathways to Understand Health Attitudes and Behaviors, Key Elements of Cultural Change to Promote Health and Well-Being

There are many examples of health-policy areas to suggest where social media may offer insight or suggest specific policy implications. For instance, the fields of infodemiology and infoveillance explore the use of social media and other Web-based data for public health, such as to predict disease outbreaks [11], explore opinion about smoking among at-risk populations [12], and investigate the impact of environmental factors such as weather on chronic pain [13]. This work has also sought to demonstrate what is possible from a methodological standpoint, such as determining the geographic distribution of Twitter users providing their location information [14], distinguishing between human Twitter users and bots (automated user accounts) [15], or case studies of social media opinion regarding specific medical conditions [16].

As these examples illustrate, the diversity of social media platforms, and of available analysis techniques, provides multiple ways in which social media track policy-relevant indicators. As outlined in Table 1 and the sections that follow, both social media data and related analytic methods contribute to this potential as a data source to understand health attitudes and behavior. For each of these areas, social media data can either complement existing health measures or provide novel ways of measuring behavior change.

Table 1. Social media data and methods for health policy action and decision making.

Social media analysis	Health policy use	Example health policy implication
Data type		
Content (text, photos, video)	Crowdsource data for public health surveillance	Use data to more efficiently inform policy interventions
Location	Build mapping and mobility patterns	Allocate resources to communities in need
Network connections	Map patterns of social relationships and interactions	Characterize social relationships and communities
Analytic method		
Content analysis	Identify health attitudes and behaviors	Build alternate measures of well-being
Network analysis	Characterize networks	Identify spread of health behaviors
Machine learning and algorithms	Predictive analytics	Monitor for early warning about disease outbreaks

Social Media Content

Policy makers who wish to estimate the impact of a proposed or newly implemented action generally rely on either new data collection or retrospective information from large-population datasets, both of which are laboriously collected and compiled before they are released. Concrete information about policy impact has to be waited for until new data are released, which may delay adjustments or follow-up efforts for months or even years. By contrast, early indications, such as indications from social media, about important health events or trends could provide policy makers with insight to inform targeted and timely interventions.

Social media postings are often spontaneous and frequent. As a result, the content of these postings is timely and can provide up-to-the-moment information. Moreover, people often post on social media from mobile devices, contributing to immediacy and, frequently, location information. Accumulating these frequent postings enables collection of large amounts of collective information that might not otherwise have been available (ie, “crowdsourcing”). Taken together, these data features may be combined with the content in powerful ways, such as for public health surveillance of disease outbreaks [9] or natural disasters [17]. Such analyses can help monitor progress of interventions or relief efforts, or improve situational awareness. A key use for social media may be to improve our understanding of the prevalence or patterns of incidents (eg, disease spread or outbreaks) that may be currently difficult to detect, measure, or quantify. Beyond discrete events, such as disasters or disease outbreaks, social media can be used to track broader trends in chronic stress, preventable hospitalizations, or global burden of disease; for instance, language use on Twitter tracks with rates of coronary heart disease [10]. However, although academic research illustrates potential for social media-based health surveillance, attempts to put this into practice have raised methodological concerns (see a following illustrative example about Google Flu Trends).

Social media may also help improve access to health information and up-to-date measures of patient and consumer experience with care. For instance, take the case of Hello Health, a small primary care practice. Hello Health doctors employ multiple social media tools, including social networks, blogs, and video chat. According to Hello Health’s doctors, such tools can improve communication between health care providers and patients, leading to increased patient engagement and satisfaction [18]. And as patients can use these social media channels to obtain information directly from their health providers, they can also seek information elsewhere online. Social media platforms provide increasingly detailed information, such as in specialized health forums (eg, WebMD, PatientsLikeMe), or on more general sites. Yelp, a Web-based review site, has partnered with ProPublica to provide additional data (eg, wait times, noise levels) to Yelp listings of health care facilities [19]. Data on these communication patterns could be instructive in understanding how to improve patient engagement, health literacy, and access to care.

The immediacy inherent in social media also suggests some potential pitfalls. Social media data are difficult to validate (eg, linking online speech to offline behaviors), and thus may be less accurate than other, more rigorously compiled datasets. In addition, social media’s rapid refresh cycle may encourage policy makers to focus on transitory quick wins or tactical improvements, to the detriment of longer-term, more strategic efforts.

Location Information

Policy decisions often depend heavily on maintaining an accurate, up-to-date picture of where residents and visitors are located. Location information may inform everyday decision making about when and where to deploy resources or disseminate public messaging. In emergencies such as natural disasters, evacuation efforts may hinge on reliable location information.

Location-based services (eg, mapping) are a key feature of many social media platforms. Consequently, location data from social media may be crucial to providing updated information about shifts in the physical environment that may be unavailable through other means. A proliferation of mapping efforts using social media-based location information (eg, by health departments), often in real time, have begun to explore geographic variations in values, health literacy, or the spread of disease outbreaks. Broad uses for these efforts could include tracking epidemiological patterns (eg, disease outbreaks, clusters, trends), or human mobility patterns that provide broader information than just health or epidemiological patterns (eg, transportation patterns, seasonal or geographic variations).

Real-time location information from social media can be used for more-accurate situational awareness of physical infrastructure, such as public transportation usage or commute times. Other possibilities to measure community investment in healthy environments include traffic data (eg, Waze), public transportation ridership, or usage of healthy alternatives, such as public bicycles. User "check-ins" on services such as Yelp and Foursquare, where users self-report their locations at specific venues such as gyms or public parks, can indicate demand for healthy places or those with opportunities for social connections. Other location-based information may be less dependent on real-time updates but no less useful. Redfin, an online real estate site, provides housing and neighborhood information, such as affordability, walkability, and safety, which may be useful in considering issues of equity. For instance, overlaying changes in home prices or median income with other location data could show the impact on a community of increasing access to health care, transportation, or healthy foods. By providing opportunities to monitor health and safety, social media and other mobile technologies could offer information about how to address stark geographic differences in life expectancy [20].

Using location information from social media may present certain difficulties. A key issue that may affect data representativeness is that only a very small proportion of social media users choose to disclose location information [14]. This may result in biased location data, particularly if there are other differences between those who choose to share location information over social media, and those who do not. There are also privacy considerations to using location data for policy making. Much can be discerned from knowing where people go, potentially including their home, workplace, or typical activities. When designing health interventions or data collection, policy makers should carefully consider how such potentially identifying characteristics are used.

Network Connections

As population data are collected, communities are often defined, out of necessity, as a simple agglomeration of characteristics made up of its individual members, without accounting for how those people may be connected. Social connections, however, can constitute a subtle form of group analysis, exploring who is similar, who talks with whom, and who may influence others' actions and attitudes. Thus, while discerning specific relationships among individuals or subgroups may be difficult, relationship and social network information can be extremely powerful.

Violence prevention efforts have compiled network information to identify and then work with individuals who have been most susceptible to perpetrating, and falling victim to, violence in their communities. Similar networks built from social media data could be used to target health messages (eg, high-risk individuals) or efficiently communicate in disaster situations (eg, notifying individuals to evacuate). For years, schools have built "telephone trees," linking parents to one another so that important information may be disseminated across the network. Social media, with its explicit relationship links, offer policy makers ready-made telephone trees across entire communities. Insight into these networks could be instrumental in how well a community responds to an emergency, providing up-to-date network maps that could reveal members who are not linked into services, and then informing policy decisions around resource allocation to fill gaps.

Social Media Content Analysis

Social media content, including message text, photos, and videos, can complement traditional attitudinal measures of health attitudes and behaviors. Insight may be extracted from any text from social media postings. For instance, attitudes regarding perceived sense of community, health interdependence, and civic engagement are linked with well-being, and currently measured by self-reported data [21]. Social media content may provide additional ways to measure perceived sense of community. Web-based conversations where people use more first-person plural pronouns (eg, "we," "us," "our") may suggest greater feelings of group cohesion or sense of community (eg, [22,23]). Alternate ways of measuring community well-being beyond economic indicators (eg, Human Development Index) are also possible. Natural language processing techniques that analyze vast amounts of text, such as those generated from social media, can be employed to develop indicators of literacy, such as creativity, language sophistication, or emergence of new forms of language (eg, slang). Similar automated techniques might also analyze nontext content, such as emoticons or emojis. Text-based measures can use analytic techniques to identify key topics of discussion, and then demonstrate varying perceptions of sense of community or civic engagement around them. This could lead to novel ways to measure civic engagement, as social media use may be predictive of voting behavior [24]. These indicators also reveal some intriguing associations with existing socioeconomic measures. For instance, a community's usage of linguistic markers of community cohesion may be related to its score on the Gini index (a measure of income distribution) [25], suggesting a possible measure of community inequality.

Social media content can improve ways of understanding health needs. Online reviews on Yelp or other sites could also serve as the basis of alternative measures of satisfaction with health providers and utilization of certain types of care (eg, complementary or alternative medicine). For example, text analysis of patient surveys have identified simple keywords, such as "excellent" or "rude," that are associated with better or worse patient experience, respectively [26]. The extent to which social media sheds light on patient experience depends on the purpose of that social media platform. Sentiment analysis of a Web-based doctor review site (RateMDs) was used to model state-level health quality statistics (eg, mortality rates, patient likelihood to seek follow-up care) [27]. In contrast, a study of tweets directed at specific hospitals showed no association between sentiment expressed in those tweets and traditional survey measures of quality of care in the hospitals [28]. This difference may be due to the fact that people writing on review sites may be motivated to provide more accurate and specific accounts of their experiences, whereas people directing comments toward health care providers may subtly alter their speech, whether it is conscious or not.

Advances in image and video recognition hold promise to capture insight from photos and videos, an ever-increasing portion of social media content. Photos and videos can show when people engage in physical activity, go outdoors, or other health behaviors. For instance, food photos, which are commonly shared, may reveal information about diet. Similarly, photos of red cups commonly used at parties could indicate problematic drinking in college students [29]. Moreover, such alcohol displays could influence others' attitudes and perceived norms about drinking behavior. With this more-granular picture, interventions could efficiently target influential individuals or groups to modify health behaviors and advance change in the cultural expectations around health and well-being.

Existing measures of subjective well-being also commonly rely on self-reported data. Complementing or replacing self-reported information with naturally occurring text or behavioral measures could improve the measurement of subjective well-being by removing certain biases (at the same time, of course, potentially introducing others). Multiple types of analyses of social media content and structure could be used for various aspects of well-being; for instance, the City of Santa Monica's Wellbeing Index analyzed Twitter text, embedded Web links, and social networks to collect sentiment and location information [30]. Perhaps the simplest would be to rely on straightforward sentiment analysis. Large-scale data on positive or negative emotions could be used to measure "happiness" (a component of overall well-being), drilled down to observe local or nongeographically-based communities, and validated against existing efforts to measure community happiness using traditional surveys (eg, Bhutan Happiness Index, Gross National Happiness, Organisation for Economic Co-operation and Development Better Life Index).

Importantly, some social media-based indicators have been linked to well-being outcomes. For instance, linguistic analysis of social media text usage of first person plural pronouns (eg, "we," "us," "our") suggests greater sense of community or group cohesion. This indicator has been linked to increased well-being; according to Schwartz and colleagues (2013): "The use of plural personal pronouns such as 'we' and 'our,' which we take to be proxies for a communal, prosocial orientation are highly correlated with the presence of LS (life satisfaction), whereas 'I' and 'my' are highly correlated with its absence." An analysis of sentiment expressed in tweets sent in various London communities found a positive relationship between sentiment and a community's socioeconomic well-being [31].

Care must be taken when drawing conclusions from what is said and shared over social media. Because social media such as tweets, Facebook posts, or Instagram photos contain information that people actively choose to share, they can provide a rich source of insight toward understanding attitudes and opinions. However, the data from this content are limited to the extent that people choose to present themselves in certain light, selectively and perhaps unconsciously adding or omitting certain content; for instance, social media information about underlying health status may depend on whether people are more likely to post about salient health issues, or health expectations that may differ by individuals. When linking to health outcomes, analyses of social media content should thus consider the context in which those words are used, rather than strictly basing conclusions on simple keyword usage [32]. Finally, the presence of duplicate, commercial, or spam accounts suggests that not all social media content is posted by individuals and therefore may not reflect health-relevant information. Although such content may not directly reflect individuals' health attitudes or behaviors, it may be relevant nonetheless, for example, tracking youth exposure to unhealthy Web-based advertising.

Network Analysis

Social media users forge both implicit and explicit connections that can help understand health interdependence, that is, the extent to which people believe their health is dependent on that of their friends or family. Network analysis of social networking platforms can help track how and whether people believe their health can influence, and be influenced by, others with whom they have social relationships. Social media ties can be either two-way (eg, Facebook friends, reciprocal Twitter mentions) or one-way (eg, Twitter follower; Facebook or Instagram likes) relationships. These directed relationships could help determine causality of associations between Web-based behaviors or suggest the directionality in which attitudes and beliefs spread.

Network analysis may be used to examine multiple aspects of social relationships, such as identifying influential people, characterizing specific communities, and the flow of information. Social media and other forms of Web-based data may also afford opportunities to analyze the implicit networks generated by participant interactions. These include, first, increased participation in the so-called "sharing-" or "gig-" economy, that is, technology-enabled service companies that facilitate exchanges between users (eg, Airbnb, Uber, TaskRabbit). Certain sharing economy platforms may be particularly amenable to this, such as home-sharing sites for travel (Airbnb, Couchsurfing) that encourage its users to meet and share experiences. One approach may be to explore partnerships with the online service companies themselves. Airbnb, for instance, actively participates in public policy issues and encourages community as part of its business, for instance, using its data to measure perceived trust and sense of community in cities [33].

Network analysis of social media may also offer new ways to track the existence and quality of cross-sector collaborations and partnerships among health organizations. Social media metadata of connections, such as Twitter mentions and followers, Facebook friends, or LinkedIn connections, can reveal structure and networks of organizational partnerships. These networks can be compared against health outcomes to show where partnerships are effective in improving well-being and also identify areas where new partnerships and collaborations would be fruitful. Network data on organizational partnerships may also explore integration of traditional community resources and health providers (eg, hospitals) with nontraditional community resources that can also influence health. Given the wide range of social media across topics and uses, social media data may also help examine how nonhealth stakeholders can play a role in improving health and well-being. For example, measuring stakeholder support for health promotion could

involve tracking social media mentions of health and well-being within a specific domain. This could involve either the communications from key organizations or other stakeholders in these sectors, as well as mentions of their support from other social media users. This could involve insight about support for workplace wellness programs, a community's attitudes and perceptions of policing efforts, or measures could track online exposure to healthy or unhealthy content (eg, advertising for unhealthy food, alcohol).

Finally, network methods may be employed not only to characterize interpersonal connections, such as among social media users, but as a way to analyze social media content. Semantic network analysis can be used to identify the co-occurrence and relationships among words. This may be useful in determining what broader linguistic patterns and concept mappings reveal about implicit health attitudes, such as whether people think of poor health as an individual responsibility or a societal failure.

Machine Learning and Algorithms

Too much data can be as much of a problem as insufficient data. When data volumes are overwhelming, it may be impossible to determine relevant data attributes and appropriate metrics. For instance, policy makers who wish to use social media data to learn about a community must first select a social media platform (eg, Twitter vs Facebook) and then determine whether to use tweet-content information, network metadata, location, images, video, or other attributes. Large amounts of data may also make it difficult or even impossible to recognize important patterns in the data, that is, to locate the signal within the noise. And as with any human endeavors, decision makers may unwittingly incorporate their own cognitive biases in interpreting the data.

Data science techniques such as sentiment analysis and machine learning can help make sense of large amounts of information (eg, combining multiple data sets) to support various forms of decision making. Sentiment analysis is a form of natural language processing that seeks to identify attitudes and emotions that are expressed in the text of, for example, social media postings. Research on Twitter data suggests that sentiment (as indicated by word use) in tweets can be used to model life satisfaction [7], happiness [6], and heart disease mortality [10] and health. Sentiment revealed in social media data can also help predict engagement in healthy behaviors, such as health insurance enrollment [8].

A well-known example in applying machine learning techniques for health predictions is Google Flu Trends. Google searches for flu and health-related terms, compared with data of flu-related doctor visits, appeared to provide early detection of influenza outbreaks, as compared with Centers for Disease Control and Prevention's existing model based on traditional data collection through the public health system [9]. Subsequent research has provided further evidence of the predictive power made possible by combining Google and Centers for Disease Control and Prevention data [34]. The Google Flu Trends example illustrates the potential value from aggregating this kind of preexisting data, and how these approaches may complement existing public health methods. However, there have been debates over the true predictive power and usefulness of using search data for public health surveillance. Suggesting caution in the use of such data, one analysis suggested that the predictive power of Google Flu Trends's model was significantly overstated [35] but could be useful in conjunction with other flu-tracking data.

Machine learning algorithms attempt to automatically classify or categorize data, such as identifying topics of discussion or objects in images. These algorithms could be used for predictive analytics that support clinical decision making, to determine pricing based on patient or community outcome data, or to contribute to personalized medicine. Social media data may also contribute to predictive analytics that aid in public health responses or planning. For instance, Yelp reviews and Twitter (see the Foodborne Chicago example) have been used to predict foodborne illness outbreaks [36-38]. Building this type of capacity could allow public health agencies to respond more quickly to unfolding public health incidents, preventing them from becoming full-blown crises. The Chicago Department of Health, which runs the Foodborne Chicago website, provides an example of an early warning system that allows people to report possible cases, and analyzes tweets for reports of food poisoning. Models with sufficient precision or specificity could even suggest imminent outbreaks, allowing authorities to respond proactively to prevent the occurrence of such incidents.

Research into Web-based indicators of trauma is common in several other fields and may be borrowed for the health-trauma context. Traumatic life events or other extremely negative experiences that occur in early childhood may be difficult to measure, either because of underreporting due to stigma or lack of ongoing measures (eg, questions about adverse childhood experiences are no longer asked as part of the Behavioral Risk Factor Surveillance System survey). As an example, a vast number of photos are shared online. Image recognition software is already employed to detect and flag traumatic events. For instance, Facebook flags what it considers questionable or inappropriate content (eg, bullying or child pornography), as well as indications of harm (eg, suicide risk) [39]. Tweets containing content indicative of suicide risk factors are correlated with actual suicide rates [40]. These techniques could be repurposed to detect instances of child abuse, sexual assault, or other adverse childhood events.

Although it may be tempting to assume that machine learning and algorithmic techniques offer impartial and equitable analyses of large volumes of data, bias nevertheless creeps into algorithms as well. This bias can involve either invalid assumptions made by those developing the algorithms, or skewed data upon which the algorithms are trained and then applied. Social media measures that do not account for algorithmic biases as well as skewed social media usage may thus inadvertently exclude or underrepresent certain population segments (eg, poor, rural) from policy making consideration. The example of Street Bump illustrates how well-intended uses of new data sources for forward-looking, informed policy decision making could have caused inadvertent harm to certain segments of the population. The City of Boston introduced a mobile phone app, Street Bump, which allowed users to report potholes they encountered. Unexpectedly, more potholes were

reported in wealthier areas than in poorer areas. Mobile phone users were presumably more affluent and also tended to drive in affluent areas. Therefore, a simple algorithm that merely allocated resources according to these results would have deepened existing inequities by widening the gap in transportation infrastructure. Poor-quality transportation infrastructure could in turn hinder emergency responders, limit access to preventive health care, or discourage social interactions and community cohesion. In contrast, high-quality transportation infrastructure could increase access for emergency response or preventive health care, encourage social interaction, and in doing so, potentially shrink equity gaps. Informed uses of such data should therefore consider how and whether methods of either data collection or analysis are representative of the populations being served, and what the potential impact on those populations may be.

Considerations for Future Social Media Analysis

As illustrated above, social media and other emerging data have the potential to contribute broadly to policies focused on health and well-being, as well as inform how social and cultural change may be underway around the importance of these issues. Policy actions intended to improve well-being must act through multiple channels, including improving communities we live in, services we use, and our attitudes toward health and well-being. Achieving a healthier future will require forward-looking methods to draw policy insight from emerging data sources, build cross-sector partnerships, and take full advantage of technological innovation. To bring this vision to reality, health and well-being policy interventions may benefit from the continued use of social media analytics, which can serve as an important complement to traditional social science data collection and analysis. Although researchers and some communities and policy makers have taken note of social media's utility, its potential is yet to be fully realized. At the same time, traditional sources of survey or administrative data have significant limitations where quantifiable, behavioral data sources such as social media, mobile devices, or other digital outputs can fill gaps. Efforts to improve quality of life should provide an opportunity to explore these areas, and encourage those conducting national and local data collection efforts to incorporate more of these emerging data sources. To illustrate, the Culture of Health Action Framework includes a measure of health-related discourse, based on Twitter data [41]. In addition, several measures of well-being and health equity use national-level survey or administrative data, which social media data could complement. For example, chronic disease burden is currently measured using disability-adjusted life years, a metric that is calculated using a variety of data sources [42]. Disease monitoring from social media data could be employed to provide a more granular picture, such as trends over time, or revealing social networks that may experience greater burden of specific chronic diseases.

Other Web-based data sources may be useful to complement information drawn from social media. Web searches and other information-seeking behaviors offer additional insight into people's concerns. In the 2014 contaminated-water crisis in Flint, Michigan, residents began searching for information about contaminated water and, as news coverage continued, broadened their searches to explore potential health impact [43]. Web search activity for health information and resources could be used to measure health attitudes in different communities. Accordingly, search data can be aggregated and then compared with other health data (eg, traditional health surveys or estimates) to build predictive models of large-scale, population-level health and health behaviors. For instance, cancer-related searches are associated with American Cancer Society estimates of cancer incidence and mortality [44].

Going forward, the value of social media data to guide policy making may rest on the ability to continually shape analyses to match the ever-shifting data sources and platforms. In other words, as the nature of social media itself changes, analysis of social media will need to continually evolve. Many important sources of data may not yet exist. Therefore, while analyzing social media data offers many advantages, several key implications should also be considered.

Limitations

The diversity of social media data sources and analytic methods suggests the need to ensure their validity: that social media data reflect real-world outcomes. Questions remain, for instance, about how to properly interpret what is shared online. What people choose to share may constitute a form of self-presentation or performance to a specific audience, rather than convey their true feelings [45]. Other data validity concerns relate to the commercial nature of social media. Profit-driven incentives may lead social media platforms to highlight certain types of sharing, or Internet service providers to prioritize certain types of Web traffic (ie, "net neutrality").

To overcome such limitations of social media data, there is a need for validation research. Validation of data and methods may include determining whether measures based on social media data track with other, more traditional measures of the same concept (eg, surveys of attitudes, public health disease monitoring). Another way to validate social media data may seek to relate it directly to real-world behaviors (eg, civic participation, health provider visits, insurance enrollment, organizational partnerships). For instance, one intriguing study showed that Google searches for mental health information follow consistent seasonal patterns, uncovering a potentially useful finding, but one which the authors noted would have to be validated against clinical or other surveillance data [46]. Although other social media research has tried to establish these relationships, much of this work is specific to a data source such as Twitter [47] or Google searches [48]. Establishing broad validity across social media platforms or analytic techniques could help move social media from research to action. Building awareness of and trust in emerging data sources increase the likelihood that they will be used to inform health policy making. Alternately, policy makers may need to accept trade-offs in using social media analytics with low validity for exploratory purposes or in surveillance and monitoring, but not necessarily when accuracy is critical, such as equitably allocating public resources.

The extent to which social media data are representative of either the general population or specific subpopulations suggests another important consideration. Care must be taken so that actions informed by nonrepresentative data do not exacerbate existing inequities by affording certain groups fewer opportunities to be heard, in effect, writing some people out of future narratives and resulting actions. For instance, compared with the general population, social media users tend to be young, educated, wealthy, and living in cities [49]. Ownership of mobile devices, from which social media content is frequently posted and accessed, is widespread but also limited to those who can afford them. People in certain geographic regions may be more likely to use particular social media platforms; for instance, homegrown services such as Weibo and WeChat are primarily used in Asia. Social media measures that do not account for these skewed demographics may thus inadvertently exclude or underrepresent certain population segments (eg, poor, rural) from analyses based on which data source or analysis method is chosen. This could significantly harm those populations if, for instance, public health resources are misallocated away from those in need, or public health interventions ineffectively target health conditions.

However, the demographics of social media users may also offer the tantalizing possibility of reaching population segments that may have been difficult to reach by traditional means of data collection [50]. For instance, samples of Twitter users may disproportionately contain individuals such as African Americans, highly educated and high-income people, and younger and urban-dwelling people [51,48]. Therefore, social media may represent a new opportunity to fully capture the voices and civic engagement of those who may have been marginalized in civic decision making. Future policy measures should seek to advance these important uses for social media and other emerging data.

Finally, one of the most widely discussed implications of social media, "big data," and other data sources is the potential impact on privacy. For instance, social media users typically do not explicitly provide consent for how their content is eventually used, unlike with traditional data collection (eg, interviews, surveys). Another consideration is that health data, like most other forms of digital data, are most commonly shared on platforms owned by private companies, which may then share that data with other entities [52]. Although individuals may assume that data shared anonymously remains as such, in fact it may be shared widely and even reidentified to infer individual identities or identify sensitive health characteristics. These privacy considerations merit further discussion to determine how to balance public good as well as benefits to individual users against these potential invasions of privacy.

Moving From Research to Action: Social Media for Policy Actions and Social Change to Advance Health and Well-Being

Social media data can form the basis of policy indicators across multiple health domains. Exploring Web-based discussion of health promotion, for instance, should provide insight into public opinion about health, and can be expected to track broad trends in attitudes and norms about health and inform interventions to shift them. Emerging social media tools and platforms could be used to characterize aspects of well-being that are not currently tracked broadly, but for which a substantial amount of research is available on how to define and measure them, such as social isolation and perceived loneliness, belongingness, toxic stress, and spirituality.

By now, there is a large and diverse body of research suggesting how to extract public health insight from social media data. Translating this knowledge into practice could begin by using social media data and analytic methods to complement existing policy data and practice, or by helping to establish relationships among diverse stakeholders who could work across sectors to advance health and well-being. First, an initial step toward using social media data in practice (eg, research, policy making) could be as a complement to more-traditional data sources that are already in use. Much of the research described in this article has compared information gleaned from social media against that of existing data, or has sought to use social media data as a way to model real-world outcomes, such as health behaviors, social connections, or spread of disease. Decision makers in either the private or public sectors could mine these techniques and topics to determine whether they may be applicable for their needs. Social media data can also be used to provide complementary evidence or analysis to support human decision making, such as individual decision making about health choices or civic decision making about resource allocation and investment strategy. Combining social media and other digital data may also help identify and suggest collaborations or partnerships that yield improved health outcomes. Possible constraints, however, include whether organizations would be willing to share sensitive or proprietary information. Analytic techniques, such as from Fiscalnote, a firm that tracks and attempts to predict legislation, could be used to help identify potential actions or policies to support cross-sector collaborations.

Second, given the commercial nature of most social media data, working with these data can produce fruitful cross-sector partnerships. A collaboration between Yelp, an online review site, and ProPublica is a good illustration of the potential for how public and private sectors can work together to provide consumers with useful health information over social media. Yelp also partnered with the cities of San Francisco and New York, with support from the White House, to provide city data of restaurant hygiene scores on Yelp business pages [53]. Part of this work included creating an open data standard that would allow other interested cities to do the same.

As social media analysis becomes more established in policy making, it will be crucial to ensure data reliability and validity. For example, developing standardized metrics could help assess effectiveness of policies based on social media data and compare policy initiatives, thus facilitating translation of knowledge (eg, infodemiology) into practice [11]. Efforts to improve data reliability could also benefit from using techniques to distinguish between insightful content generated by human social media users and automated posts from bots [15].

Future efforts should also carefully consider how to preserve civil liberties, such as whether machine-learning algorithms

introduce inadvertent biases, and how to deal with privacy concerns, such as how health information or personally identifiable information may be used. As described earlier, this is particularly important because a great deal of online health information is held by private corporations [52]. Libert also points out the possibility that people are treated differently or discriminated against on the basis of perceived health information about them, whether or not it is accurate. The implications of how these data are used, either by the companies who hold it or to whom they provide it, warrant further consideration. For instance, what are the conditions under which private corporations will provide or allow access to these data? Should policy makers seek to incentivize partnerships or to prevent issues inherent in-data mining, such as inadvertent discrimination or privacy invasions? Similar privacy concerns also arise in the context of how governments may choose to use health-related insights generated from social media data. Access to vast amounts of social media data could inform policy making to improve health and well-being but could also be misused in ways that undermine health privacy and confidentiality.

In summary, a number of potential approaches could improve the accessibility and utility of social media in policy making. Given that much of these data are held commercially, the vast market opportunities that companies continue to envision with digital health data, as well as increased corporate interest in well-being and social impact, opportunities should exist to partner broadly with stakeholders across a range of sectors. Accordingly, increased global interest in well-being measurement for policy making affords additional opportunities to collaborate and build data analysis capabilities. Taken together, social media and other emerging data sources provide multiple avenues to help track and motivate wide-ranging and truly inclusive policy action to improve health and well-being for all.

Acknowledgments

The author thanks Carolyn Miller and Matthew Trujillo of the Robert Wood Johnson Foundation, and Anita Chandra, Katherine Carman, Laurie Martin, and Chris Nelson of the RAND Corporation, who provided input on earlier drafts of this manuscript. Sarah Fieldhouse provided administrative support. This work was supported by the Robert Wood Johnson Foundation.

Conflicts of Interest

A member of the author's immediate family is an employee at Google.

References

1. Plough AL. Building a culture of health: a critical role for public health services and systems research. Am J Public Health 2015 Apr;105 Suppl 2:S150-S152. [doi: 10.2105/AJPH.2014.302410] [Medline: 25689198]
2. Yeung D, May Warren L, Trujillo M. The Health Care Blog. 2017 Jan 25. What 32 million tweets tell us about health & the Twitterverse URL: http://thehealthcareblog.com/blog/2017/01/25/what-32-million-tweets-tell-us-about-health/ [accessed 2017-03-27] [WebCite Cache ID 6pd1k5TiZ]
3. Robert Wood Johnson Foundation. Culture of Health. 2017. Making health a shared value URL: http://www.cultureofhealth.org/en/taking-action/making-health-a-shared-value.html [accessed 2017-06-26] [WebCite Cache ID 6rW4sXZi9]
4. Smith A. Pew Research Center. 2015 Apr. U.S. Smartphone Use in 2015 URL: http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/ [WebCite Cache ID 6xNqD2Mlz]
5. Chung CK, Jones C, Liu A, Pennebaker JW. Predicting success and failure in weight loss blogs through natural language use. 2008 Presented at: International Conference on Weblogs and Social Media; March 30-April 2, 2008; Seattle, WA.
6. Mitchell L, Frank MR, Harris KD, Dodds PS, Danforth CM. The geography of happiness: connecting Twitter sentiment and expression, demographics, and objective characteristics of place. PLoS One 2013;8(5):e64417 [FREE Full text] [doi: 10.1371/journal.pone.0064417] [Medline: 23734200]
7. Schwartz HA, Eichstaedt JC, Kern ML, Dziurzynski L, Agrawal M, Park GJ, et al. Characterizing geographic variation in wellbeing using Tweets. 2013 Presented at: International Conference on Weblogs and Social Media; 2013; Boston, MA.
8. Wong CA, Sap M, Schwartz A, Town R, Baker T, Ungar L, et al. Twitter sentiment predicts Affordable Care Act marketplace enrollment. J Med Internet Res 2015 Feb 23;17(2):e51 [FREE Full text] [doi: 10.2196/jmir.3812] [Medline: 25707038]
9. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
10. Eichstaedt JC, Schwartz HA, Kern ML, Park G, Labarthe DR, Merchant RM, et al. Psychological language on Twitter predicts county-level heart disease mortality. Psychol Sci 2015 Feb;26(2):159-169 [FREE Full text] [doi: 10.1177/0956797614557867] [Medline: 25605707]
11. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]

12. Rose SW, Jo CL, Binns S, Buenger M, Emery S, Ribisl KM. Perceptions of menthol cigarettes among Twitter users: content and sentiment analysis. J Med Internet Res 2017 Feb 27;19(2):e56 [FREE Full text] [doi: 10.2196/jmir.5694] [Medline: 28242592]
13. Haghighi PD, Kang YB, Buchbinder R, Burstein F, Whittle S. Investigating subjective experience and the influence of weather among individuals with fibromyalgia: a content analysis of Twitter. JMIR Public Health Surveill 2017 Jan 19;3(1):e4 [FREE Full text] [doi: 10.2196/publichealth.6344]
14. Burton SH, Tanner KW, Giraud-Carrier CG, West JH, Barnes M. "Right time, right place" health communication on Twitter: value and accuracy of location information. J Med Internet Res 2012 Nov 15;14(6):e156 [FREE Full text] [doi: 10.2196/jmir.2121] [Medline: 23154246]
15. Allem JP, Ferrara E, Uppu SP, Cruz TB, Unger JB. E-cigarette surveillance with social media data: social bots, emerging topics, and trends. JMIR Public Health Surveill 2017 Dec 20;3(4):e98 [FREE Full text] [doi: 10.2196/publichealth.8641] [Medline: 29263018]
16. Nascimento TD, DosSantos MF, Danciu T, DeBoer M, van Holsbeeck H, Lucas SR, UMSoD (Under)Graduate Class Of 2014, et al. Real-time sharing and expression of migraine headache suffering on Twitter: a cross-sectional infodemiology study. J Med Internet Res 2014 Apr;16(4):e96 [FREE Full text] [doi: 10.2196/jmir.3265] [Medline: 24698747]
17. Kryvasheyeu Y, Chen H, Moro E, Van Hentenryck P, Cebrian M. Performance of social network sensors during Hurricane Sandy. PLoS One 2015;10(2):e0117288 [FREE Full text] [doi: 10.1371/journal.pone.0117288] [Medline: 25692690]
18. Hawn C. Take two aspirin and tweet me in the morning: how Twitter, Facebook, and other social media are reshaping health care. Health Aff (Millwood) 2009;28(2):361-368 [FREE Full text] [doi: 10.1377/hlthaff.28.2.361] [Medline: 19275991]
19. Stoppelman J. Yelpblog. 2015 Aug 05. Yelp's consumer protection initiative: ProPublica partnership brings medical info to Yelp URL: https://www.yelpblog.com/2015/08/yelps-consumer-protection-initiative-propublica-partnership-brings-medical-info-to-yelp [accessed 2018-02-20] [WebCite Cache ID 6xNriZTX0]
20. Graham G, Ostrowski ML, Sabina A. Health Affairs. 2015 Aug 06. Defeating the ZIP code health paradigm: data, technology, and collaboration are key URL: http://healthaffairs.org/blog/2015/08/06/defeating-the-zip-code-health-paradigm-data-technology-and-collaboration-are-key/ [accessed 2018-02-20] [WebCite Cache ID 6pd1ryfKe]
21. Chandra A, Miller CE, Acosta JD, Weilant S, Trujillo M, Plough A. Drivers of health as a shared value: mindset, expectations, sense of community, and civic engagement. Health Aff (Millwood) 2016 Nov 01;35(11):1959-1963. [doi: 10.1377/hlthaff.2016.0603] [Medline: 27834233]
22. Pennebaker JW, Booth RJ, Francis ME. Citeseerx.ist.psu. Austin, TX: LIWC.net; 2007. Operator's manual: linguistic inquiry and word count: LIWC2007 URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.600.7227&rep=rep1&type=pdf [accessed 2018-02-20]
23. Pennebaker JW, Chung CK, Ireland M, Gonzales A, Booth RJ. LIWC.net. 2007. The development and psychometric properties of LIWC URL: http://www.liwc.net/LIWC2007LanguageManual.pdf [accessed 2018-02-20] [WebCite Cache ID 6xNsHWmzi]
24. DiGrazia J. Using Internet search data to produce state-level measures: the case of tea party mobilization. Socio Meth Res 2015 Oct 14;46(4):898-925. [doi: 10.1177/0049124115610348]
25. Pennebaker JW. The Secret Life of Pronouns: What Our Words Say About Us. New York: Bloomsbury Press; 2011.
26. Maramba ID, Davey A, Elliott MN, Roberts M, Roland M, Brown F, et al. Web-based textual analysis of free-text patient experience comments from a survey in primary care. JMIR Med Inform 2015 May 06;3(2):e20 [FREE Full text] [doi: 10.2196/medinform.3783] [Medline: 25947632]
27. Wallace BC, Paul MJ, Sarkar U, Trikalinos TA, Dredze M. A large-scale quantitative analysis of latent factors and sentiment in online doctor reviews. J Am Med Inform Assoc 2014;21(6):1098-1103 [FREE Full text] [doi: 10.1136/amiajnl-2014-002711] [Medline: 24918109]
28. Greaves F, Laverty AA, Cano DR, Moilanen K, Pulman S, Darzi A, et al. Tweets about hospital quality: a mixed methods study. BMJ Qual Saf 2014 Oct;23(10):838-846 [FREE Full text] [doi: 10.1136/bmjqs-2014-002875] [Medline: 24748372]
29. Moreno MA, D'Angelo J, Kacvinsky LE, Kerr B, Zhang C, Eickhoff J. Emergence and predictors of alcohol reference displays on Facebook during the first year of college. Comput Human Behav 2014 Jan;30:87-94 [FREE Full text] [doi: 10.1016/j.chb.2013.07.060] [Medline: 24415846]
30. The City of Santa Monica Wellbeing Project. The Wellbeing Project. 2015 Apr. Creating a city for wellbeing: key findings about wellbeing perspectives and assets in Santa Monica URL: https://wellbeing.smgov.net/Media/Default/docs/WellbeingProject-IndexFindings-April2015.pdf [accessed 2020-02-20] [WebCite Cache ID 6ppB1IDrX]
31. Quercia D, Ellis J, Capra L, Crowcroft J. Tracking "Gross Community Happiness" from Tweets. 2012 Presented at: Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work; February 11-15, 2012; Seattle, WA. [doi: 10.1145/2145204.2145347]
32. Mowery D, Smith H, Cheney T, Stoddard G, Coppersmith G, Bryan C, et al. Understanding depressive symptoms and psychosocial stressors on Twitter: a corpus-based study. J Med Internet Res 2017 Feb 28;19(2):e48 [FREE Full text] [doi: 10.2196/jmir.6895] [Medline: 28246066]

33. Newman R, Antin J. Medium. 2016 Mar 14. Building for trust URL: https://medium.com/airbnb-engineering/building-for-trust-503e9872bbbb [accessed 2017-04-17] [WebCite Cache ID 6pis9U2mK]
34. Davidson MW, Haim DA, Radin JM. Using networks to combine "big data" and traditional surveillance to improve influenza predictions. Sci Rep 2015 Jan 29;5:8154 [FREE Full text] [doi: 10.1038/srep08154] [Medline: 25634021]
35. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. Science 2014 Mar 14;343(6176):1203-1205. [doi: 10.1126/science.1248506] [Medline: 24626916]
36. Harrison C, Jorder M, Stern H, Stavinsky F, Reddy V, Hanson H, Centers for Disease ControlPrevention (CDC). Using online reviews by restaurant patrons to identify unreported cases of foodborne illness - New York City, 2012-2013. MMWR Morb Mortal Wkly Rep 2014 May 23;63(20):441-445 [FREE Full text] [Medline: 24848215]
37. Nsoesie EO, Kluberg SA, Brownstein JS. Online reports of foodborne illness capture foods implicated in official foodborne outbreak reports. Prev Med 2014 Oct;67:264-269 [FREE Full text] [doi: 10.1016/j.ypmed.2014.08.003] [Medline: 25124281]
38. Schomberg JP, Haimson OL, Hayes GR, Anton-Culver H. Supplementing public health inspection via social media. PLoS One 2016;11(3):e0152117 [FREE Full text] [doi: 10.1371/journal.pone.0152117] [Medline: 27023681]
39. Davis PK, Perry WL, Brown RA, Yeung D, Roshan P, Voorhees P. Using Behavioral Indicators to Help Detect Potential Violent Acts: A Review of the Science Base. Santa Monica, CA: RAND Corporation; 2013. URL: https://www.rand.org/content/dam/rand/pubs/research_reports/RR200/RR215/RAND_RR215.pdf [WebCite Cache ID 6xt8F2v6B]
40. 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]
41. Chandra A, Acosta JD, Carman KG, Dubowitz T, Leviton L, Martin LT, et al. Building a National Culture of Health: Background, Action Framework, Measures, and Next Steps. Santa Monica, CA: RAND Corporation; 2016. URL: https://www.rand.org/content/dam/rand/pubs/research_reports/RR1100/RR1199/RAND_RR1199.pdf [WebCite Cache ID 6rW6V62MN]
42. Murray CJ. Quantifying the burden of disease: the technical basis for disability-adjusted life years. Bull World Health Organ 1994;72(3):429-445 [FREE Full text] [Medline: 8062401]
43. Matsa KE, Mitchell A, Stocking G. Journalism. Searching for news: the Flint water crisis URL: http://www.journalism.org/essay/searching-for-news/ [accessed 2017-06-26] [WebCite Cache ID 6rW5GezYD]
44. Cooper CP, Mallon KP, Leadbetter S, Pollack LA, Peipins LA. Cancer Internet search activity on a major search engine, United States 2001-2003. J Med Internet Res 2005 Jul 1;7(3):e36 [FREE Full text] [doi: 10.2196/jmir.7.3.e36] [Medline: 15998627]
45. Marwick AE, Boyd D. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. New Media Soc 2010 Jul 07;13(1):114-133. [doi: 10.1177/1461444810365313]
46. Ayers JW, Althouse BM, Allem JP, Rosenquist JN, Ford DE. Seasonality in seeking mental health information on Google. Am J Prev Med 2013 May;44(5):520-525. [doi: 10.1016/j.amepre.2013.01.012] [Medline: 23597817]
47. Paul MJ, Dredze M. You are what you tweet: analyzing twitter for public health. 2011 Presented at: Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media; 2011; Menlo Park, California URL: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2880/3264
48. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The use of google trends in health care research: a systematic review. PLoS One 2014;9(10):e109583 [FREE Full text] [doi: 10.1371/journal.pone.0109583] [Medline: 25337815]
49. Perrin A. Pew Research Center. 2015 Oct. Social media usage 2005-2015 URL: http://www.pewinternet.org/2015/10/08/social-networking-usage-2005-2015 [WebCite Cache ID 6pitAG1wZ]
50. Kim Y, Huang J, Emery S. The research topic defines "noise" in social media data - a response from the authors. J Med Internet Res 2017 Jun 02;19(6):e165 [FREE Full text] [doi: 10.2196/jmir.6824] [Medline: 28576756]
51. Duggan M, Ellison NB, Lampe C, Lenhart A, Madden M, Rainie L, et al. Pew Research Center. 2015 Jan. Social media update 2014 URL: http://www.pewinternet.org/2015/01/09/social-media-update-2014/ [WebCite Cache ID 6xNtP90VM]
52. Libert T. Privacy implications of health information seeking on the web. Commun ACM 2015 Feb 23;58(3):68-77. [doi: 10.1145/2658983]
53. Stoppelman J. Yelpblog. 2013. Now on Yelp: restaurant inspection scores URL: https://www.yelpblog.com/2013/01/introducing-lives [accessed 2017-04-14] [WebCite Cache ID 6pisO48yd]

Abbreviations

GDP: gross domestic product

Edited by G Eysenbach; submitted 18.07.17; peer-reviewed by J Meng, M Bardus, A Benetoli; comments to author 23.11.17; revised version received 17.01.18; accepted 23.01.18; published 19.03.18.

Please cite as:
Yeung D
Social Media as a Catalyst for Policy Action and Social Change for Health and Well-Being: Viewpoint
J Med Internet Res 2018;20(3):e94
URL: http://www.jmir.org/2018/3/e94/
doi:10.2196/jmir.8508
PMID:29555624

Viewpoint

Lessons Learned From a Living Lab on the Broad Adoption of eHealth in Primary Health Care

Ilse Catharina Sophia Swinkels[1,2], PhD; Martine Wilhelmina Johanna Huygens[2,3], PhD; Tim M Schoenmakers[1,2,4], PhD; Wendy Oude Nijeweme-D'Hollosy[2,5], PhD; Lex van Velsen[2,5,6], PhD; Joan Vermeulen[2,3,7], PhD; Marian Schoone-Harmsen[2,8], MSc; Yvonne JFM Jansen[2,8,9], PhD; Onno CP van Schayck[2,10], MD, PhD; Roland Friele[1,2,11], PhD; Luc de Witte[2,3,12], MD, PhD

[1]Netherlands Institute for Health Services Research, Utrecht, Netherlands
[2]Centre for Care Technology Research, Maastricht, Netherlands
[3]Department of Health Services Research, School for Public Health and Primary Care, Maastricht University, Maastricht, Netherlands
[4]Department of Clinical Psychology, University of Amsterdam, Amsterdam, Netherlands
[5]Telemedicine, Biomedical Signals and Systems, University of Twente, Enschede, Netherlands
[6]Roessingh Research & Development, Enschede, Netherlands
[7]Lunet Zorg, Eindhoven, Netherlands
[8]Work Health Technology Expertise Group, Netherlands Organisation for Applied Scientific Research TNO, Leiden, Netherlands
[9]Robuust, Eindhoven, Netherlands
[10]Department of Family Medicine, School for Public Health and Primary Care, Maastricht University, Maastricht, Netherlands
[11]Tranzo, Tilburg School of Social and Behavorial Sciences, Tilburg University, Tilburg, Netherlands
[12]Research Center Technology and Care, Zuyd University of Applied Sciences, Heerlen, Netherlands

Corresponding Author:
Ilse Catharina Sophia Swinkels, PhD
Netherlands Institute for Health Services Research
PO Box 1568
Utrecht, 3500 BN
Netherlands
Phone: 31 302729771
Email: i.swinkels@nivel.nl

Abstract

Background: Electronic health (eHealth) solutions are considered to relieve current and future pressure on the sustainability of primary health care systems. However, evidence of the effectiveness of eHealth in daily practice is missing. Furthermore, eHealth solutions are often not implemented structurally after a pilot phase, even if successful during this phase. Although many studies on barriers and facilitators were published in recent years, eHealth implementation still progresses only slowly. To further unravel the slow implementation process in primary health care and accelerate the implementation of eHealth, a 3-year Living Lab project was set up. In the Living Lab, called eLabEL, patients, health care professionals, small- and medium-sized enterprises (SMEs), and research institutes collaborated to select and integrate fully mature eHealth technologies for implementation in primary health care. Seven primary health care centers, 10 SMEs, and 4 research institutes participated.

Objective: This viewpoint paper aims to show the process of adoption of eHealth in primary care from the perspective of different stakeholders in a qualitative way. We provide a real-world view on how such a process occurs, including successes and failures related to the different perspectives.

Methods: Reflective and process-based notes from all meetings of the project partners, interview data, and data of focus groups were analyzed systematically using four theoretical models to study the adoption of eHealth in primary care.

Results: The results showed that large-scale implementation of eHealth depends on the efforts of and interaction and collaboration among 4 groups of stakeholders: patients, health care professionals, SMEs, and those responsible for health care policy (health care insurers and policy makers). These stakeholders are all acting within their own contexts and with their own values and expectations. We experienced that patients reported expected benefits regarding the use of eHealth for self-management purposes, and health care professionals stressed the potential benefits of eHealth and were interested in using eHealth to distinguish

themselves from other care organizations. In addition, eHealth entrepreneurs valued the collaboration among SMEs as they were not big enough to enter the health care market on their own and valued the collaboration with research institutes. Furthermore, health care insurers and policy makers shared the ambition and need for the development and implementation of an integrated eHealth infrastructure.

Conclusions: For optimal and sustainable use of eHealth, patients should be actively involved, primary health care professionals need to be reinforced in their management, entrepreneurs should work closely with health care professionals and patients, and the government needs to focus on new health care models stimulating innovations. Only when all these parties act together, starting in local communities with a small range of eHealth tools, the potential of eHealth will be enforced.

(J Med Internet Res 2018;20(3):e83) doi:10.2196/jmir.9110

KEYWORDS

telemedicine; primary health care; implementation; patient involvement; entrepreneurship; health personnel; policy makers

Introduction

Needs for Real-World View on eHealth Adoption

Electronic health (eHealth) solutions are expected to empower patients and maintain or improve health outcomes, while generating cost-effective gains and lowering primary health care professionals' workload [1,2]. However, it appears to be difficult to embed them in daily health care routines [3]. Often, use of eHealth services stops when research projects are finished, even when successful. Moreover, there is still uncertainty about their effectiveness in daily practice [2,4,5]. Therefore, the success rate to date of eHealth in primary health care is low [6]. Current evidence on eHealth and care technologies is mainly based on clinical trials and isolated eHealth applications. Van Gemert-Pijnen et al (2011) suggest that evaluations should not focus exclusively on measuring outcome variables (via randomized controlled trials) but should also include in-depth process data concerning the usage of eHealth [7].

It is suggested that successful implementation of eHealth asks for a complex innovation approach [6]. Numerous factors are related to its success, including characteristics of the end users, the function and usability of the intervention, the technical infrastructure, change management of health care organizations, the health care system, and financial business models [2,4,6,7]. It can be stated that 4 groups of stakeholders are responsible for a successful implementation of eHealth solutions: patients, health care professionals, entrepreneurs, and those responsible for health care policy (policy makers and health care insurers) [8]. Eysenbach (2001) stated in 2001 that eHealth is an emerging field at the intersection of medical informatics, public health, and businesses [9]. However, literature combining the views of these different fields and describing their challenges systematically is scarce.

With this paper, we aim to fill this gap and will describe the challenges that arose when patients, health care professionals, and entrepreneurs collaborated in a Living Lab setting to select, integrate, implement, and evaluate eHealth in primary health care. Hereby, our aim is not to test the relationships and interactions between different factors and stakeholders. Rather, we aim to show, in a qualitative way, the process of adoption of eHealth in primary care from the perspective of different stakeholders. This provides a real-world view on how such a process occurs, including successes and failures related to the different perspectives. Our paper serves as an illustration that underlines the importance of including all 4 stakeholders, having a shared vision statement, and enabling all partners to invest time or money, as only then can the expected potential of eHealth solutions be reached. After providing the rationale for our Living Lab project and a short description of our methods, we reflect upon our findings in 4 sections—patients as stakeholders; health care professionals as stakeholders; entrepreneurs as stakeholders; and health care insurers and policy makers as stakeholders. On the basis of these findings, we have been able to develop lessons learned, which seem to be important in positively shaping the outcome of eHealth implementation and adoption in future primary health care.

Rationale for the Living Lab Project "eLabEL"

In 2012-2013, when writing the grant proposal for the eLabEL project, it was already known that much of the eHealth technology being developed did not reach primary care practice because of a suboptimal fit between the needs in primary care and the technical solutions [10]. Simultaneously, there was very little knowledge about what it takes to bring such technologies into practice.

With eLabEL, we aimed to contribute to this knowledge, and to bridge technology and implementation. We believed, and still believe, that incorporating eHealth into daily practice is essential for optimal effects on quality and efficiency of health care. In other words, traditional health care should change to "technology-supported health care." For such a change, not only a technological innovation but also a societal innovation is essential. Furthermore, according to Van Velsen and colleagues (2013), a multidisciplinary development approach is necessary [11]. Van Gemert-Pijnen et al (2011) stated that relevant stakeholders should collaborate, and research should consist of qualitative and quantitative elements [7].

From this perspective, we, as researchers, started the eLabEL project in 2013, together with entrepreneurs, patients, and health care professionals. eLabEL was aimed at establishing a Living Lab in which patients, health care professionals, entrepreneurs, and researchers could collaborate during the selection, integration, implementation, and evaluation of mature eHealth-tools in primary health care [12]. According to the European Network of Living Labs, we defined a Living Lab as a user-centered, open innovation ecosystem based on a

systematic user cocreation approach, integrating research and innovation processes in real-life communities and settings [13].

In this project, the focus was on two types of mature eHealth technologies: (1) online communication services which can be used by all patients in the practice and (2) eHealth for self-management purposes for those with a chronic somatic condition. Textbox 1 provides a description of the eLabEL project, and Table 1 provides a description of the characteristics of the participating primary health care centers.

Textbox 1. The eLabEL project as illustration.

The eLabEL project was conducted from September 2013 until December 2016 in the Netherlands. We aimed at the establishment of Dutch Living Labs in which integrated eHealth applications would become part of regular health care. Concurrently, we aimed to study the consequences of the integration of eHealth applications in primary care, as well as technical barriers and facilitators.

Seven primary health care centers participated in eLabEL. These were recruited via the network of the participating research partners or positively responded to the recruitment call that was published in a press release and at the project's website. In these centers, at least one general practitioner, physical therapist, practice nurse, and nurse assistant provided health care to the community. Participating practices varied in type of organization, experiences with eHealth, patients' characteristics, and region. Characteristics of these health care centers can be found in Table 1.

Patients of these primary health care centers were also invited to participate. Ten enterprises participated in the Living Labs. These were mainly small- and medium-sized enterprises (SMEs) and offered different eHealth applications or services, varying from videoconferencing and online coaches for patients with chronic diseases to activity sensors and data warehousing. These SMEs were recruited via the network of the participating research partners. Some of them already participated in prior research projects. Also 4 research institutes, collaborating in the Centre for Care Technology Research, participated. These profit and nonprofit organizations collaborated to select and integrate mature eHealth technologies for implementation in primary care. One or two members of each research institute coordinated the project.

In the Living Lab patients, health care professionals, entrepreneurs, and researchers were invited to have close contact with each other during the whole project. In practice, the following activities took place:

- At the start of the project, needs and expectations of patients and health care professionals were inventoried via focus groups and interviews. These needs were linked to existing eHealth applications developed by the SMEs.
- Regular group sessions were held with the SMEs in which they discussed integrating technology and explored a viable business model.
- Two group sessions were held with health care professionals from all participating centers.
- Two group sessions were held with health care professionals and entrepreneurs.
- Regular meetings were held with the individual practices.
- Meetings were held with policy makers and health care insurers.

The final eLabEL package exists of the following eHealth applications:

- A service to provide online video consultations
- An online self-management coach for people with chronic obstructive pulmonary disease (COPD), which supports them into a healthier lifestyle, monitors their health status, and signals decline of health status
- An online coach for patients under treatment by the physiotherapist to support them in doing exercises at home by giving online training schemes and videos
- An application developed to coordinate multidisciplinary care around a patient. In this application, patients were able to add health care professionals, family, or other caregivers and those persons could read and share information

These applications were integrated in one infrastructure with single sign on for patients.

Table 1. Characteristics of the participating primary health care centers. Ca: circa.

Number	Organization	Number of patients (2013)	Region of the Netherlands	Remarks
1	Health care center	Ca 13,000	Mid	
2	Health care center	Ca 8000	South	Located in deprived urban area
3	General practice	Ca 3500	South	
4	Health care center	Ca 5000	West	Located in deprived urban area
5	General practice	Ca 5500	North	Patients mainly students
6	Health care center	Ca 14,500	Mid	
7	General practice	Ca 6500	North	

Methods

A qualitative design was used to study the processes of adoption of eHealth in our project. We systematically analyzed all reflective and process-based notes from meetings with health care professionals, the scientific project members, members of the management team of the Centre for Care Technology research, members of societal organizations, health care insurers, and enterprises. Furthermore, data from interviews and focus groups on the needs and expectations of health care professionals and patients were included in the analyses, as well as interviews on adoption and implementation of eHealth.

In total, 30 patients with a chronic disease, that is, diabetes, chronic obstructive pulmonary disease (COPD), or a cardiovascular condition, participated in 5 semistructured focus group interviews in the first year of the project. Those patients were recruited in 4 primary care centers by the health care professional. Mean age was 68 years and 73% (22/30) were male. In these focus groups, the following themes were discussed: (1) the impact of the chronic disease on patients' daily life, (2) their opinions and needs regarding self-management, and (3) their expectations and needs regarding, and willingness to use, eHealth for self-management purposes. See Huygens et al (2016) for a detailed description of the focus group method [14]. In addition, 30 health care professionals (9 general practitioners, 8 physical therapists, 8 nurse practitioners, and 5 supporting staff members from the eLabEL practices) were interviewed in the first months of the project. Themes discussed in these semistructured interviews were (1) the centers' technical infrastructure, (2) positive and negative work-related experiences with information technology, and (3) future expectations and needs of eHealth. See Oude Nijeweme-d'Hollosy et al (2015) for a detailed description of the interview method [15]. Eight care managers from the eLabEL-practices were interviewed in 2016. In these interviews, the expected facilitating and inhibiting factors for adoption and implementation of the eHealth tools were discussed.

For the analyses, we used a coding scheme based on four theoretical models to initially structure our findings. Wagemakers' model (2010) focuses on collaboration among multidisciplinary organizations in health care [16]. Nystrom's model (2014) was used because of its focus on different role approaches within a collaboration [17]. The model of Geels (2002) describes new technologies as arising and maturing within existing technology systems [18]. Fleuren et al (2004) state that the success rate of an innovation is dependent on the level of the innovation itself, end users, organization, and the social-political context [19]. All elements in these models are included in the coding scheme. More information about these theoretical models can be found in Multimedia Appendix 1.

To shed light on the process of implementation, we performed a qualitative summative process evaluation, in which analyses were performed 21 months after the start of the project and at the end of it. At 21 months after the start, documents were allocated among some of the authors (IS, MH, WH, LV, JV, MS, and YJ). Each set of documents was coded using the coding scheme and then thematically summarized by the author. Each summary was then checked by 1 researcher of another research institute. At the end, for pragmatic reasons, 1 researcher (MH) coded and summarized the last set of documents, and 4 researchers checked the summary (IS, WH, LV, and MS). A summary of the findings of both rounds of analyses was shared with the SMEs and health care professionals for their confirmation (member check procedure). Furthermore, YJ has observed the project as action researcher. On the basis of these procedures, the findings were used to describe the process of implementation from the perspectives of the identified stakeholders.

Results

Patients

Envisaged Role in eLabEL

eLabEL aimed at a user-centred design. Patients were intended to be actively involved in the selection and implementation of eHealth solutions. With actively involved, we mean that their input is collected and used from the start (selection phase) to the end (implementation). This way, we expected to stimulate the use of eHealth by patients. However, active patient involvement was only achieved to a minor extent. In addition, we found that patient involvement does not always guarantee usage of specific eHealth technologies on a broad scale because not every patient seems willing to use eHealth.

Patient Involvement in Research

The first way to involve patients was by organizing focus groups to investigate their expectations and needs regarding eHealth. Patients had to be recruited by health care professionals to participate in these group interviews. However, it was difficult for them to encourage patients to participate. According to the health care professionals, one of the main reasons was that patients were tired of participating in research. Therefore, organizing patient involvement was more time-consuming than expected. In addition to the focus group interviews, we attempted to set up a patient panel for the active involvement of patients during the entire project. However, this resulted in only a few positive responses. We were more hesitant to encourage health care professionals to recruit more patients for this panel, as the first study already required significant effort. Furthermore, throughout the project, actively involving patients to incorporate the patient perspective in the project became of secondary importance. The focus of eLabEL shifted toward the development of an integrated eHealth structure and the investigation of barriers for its slow development and implementation. As a result, health care professionals did not offer it to their patients.

Willingness to Use eHealth Differs Between Patients

Despite the difficulties in involving patients, we did organize 5 focus groups with patients with a chronic condition. Detailed results from these focus groups about self-management and use of eHealth are published by Huygens et al [14]. Briefly, it showed that patients reported expected benefits regarding the use of eHealth for self-management purposes. For example, a patient with diabetes reported:

> *If you can monitor automatically, you get customised care more quickly. Currently, you're going to the care practice 4 times a year, and in the period in between you stay at the same value [of insulin], while you maybe should have changed it in the meantime, but you didn't know that.* [Focus groups, patient with diabetes]

However, many patients also did not feel a need to use eHealth for self-management purposes. It seemed that the perceived benefits of using eHealth should outweigh the negative consequences of frequently having to take action to deal with the disease, which reminds patients about having a disease. A patient with a cardiovascular condition that had little impact on his daily life mentioned the following:

> *The disadvantage is that I'm feeling more like a patient [because of frequent monitoring]: man suffers most from the suffering he fears.* [Focus groups, patient with a cardiovascular condition]

Health Care Professionals

Envisaged Role in eLabEL

The role of the health care professionals was to actively participate in the Living Lab settings. They were expected to provide input regarding their own needs and requirements regarding eHealth and its implementation. In addition, they had to use the applications in their daily care processes and encourage and support their patients to use them. Our intention was that health care professionals would implement and use eHealth without the help of the research team. However, we found that the organization of primary health care was inadequate and not sufficiently equipped for doing so as we explain in the upcoming section.

Health Care Professionals See Potential in eHealth

The participating health care professionals stressed the potential benefits of eHealth. Professionals identified the rising development of eHealth technologies, the emergence of different eHealth initiatives, and their opportunities for better health care. In addition, care professionals indicated that they were interested in using eHealth to distinguish themselves from other care organizations. Providing extramural care, monitoring patients at a distance, empowering and supporting self-management of patients, providing more intensive care in less time, providing care during out-of-office hours, and increasing the quality of care, were frequently mentioned anticipated benefits of eHealth. Health care professionals believed that by using eHealth for people with mild conditions, they could save time and provide extra time to those with more severe conditions.

Support for Incorporating eHealth in Daily Practice

After deciding which eHealth technologies they wanted to use, it was not just a matter of connecting the technology. We experienced many difficulties in the implementation of eHealth in the care practices. First, health care professionals needed support for eHealth usage, including clear instruction material, a helpdesk, and, most importantly, time to gain experience with eHealth, as they had not worked with the selected eHealth applications previously. In addition, for health care professionals it was unclear how eHealth could be successfully integrated into their daily work. Workflow, responsibilities, and roles needed to change, and they did not know how to approach this. Moreover, eHealth was not integrated into the electronic medical records or protocols. This made it difficult for the health care professionals to imagine how to integrate eHealth into their daily care processes. Furthermore, health care professionals expected and experienced problems regarding motivating patients to use eHealth. Clear instruction material and tips (eg, from other care professionals) to encourage and convince patients to use eHealth were needed. In addition, health care professionals indicated that they did not want to innovate without the help and encouragement of other health care professionals within and outside their own organization. It appeared that the innovation should fit with the ambitions and plans of the local care community.

Convincing Partners Within and Outside the Practice

So, health care professionals needed support on different levels during the implementation, more than we expected. For these support activities funding was needed, which was not covered by the budget for the project. Several care practices tried to apply for eHealth funding. However, we experienced that it was complex for them to organize this. Often, they lacked knowledge, expertise, or resources to apply for eHealth funding. Professionals mentioned that in the current financial model, they had to pay the costs (time and money) for eHealth implementation, while the health insurer would receive the proposed benefits in terms of cost reductions (also known as the wrong pocket problem).

In addition, health care professionals already experienced a high time pressure in regular care processes and in keeping up with bureaucratic and legal changes, resulting in a lack of time to adopt eHealth. Moreover, in most practices, eHealth was not mentioned in vision and mission statements. Furthermore, the care professionals and managers who agreed to participate in the eLabEL project were not the ones that actually had to work with the applications in real practice. An "eHealth-minded" care manager does not guarantee the actual use of eHealth by his or her colleagues when there is no clear vision on eHealth in the care organization or space for innovation. The aforementioned reasons resulted in low priority for eHealth implementation. As summarized by one of the managers:

> *I am supporter of such innovations in health care, but I also see that they conflict with every day practice. General practitioners are up to their ears in work. They have no time for implementation. Primary health care professionals experience extreme pressure due to the substitution from secondary to primary care, which is bothering them. Besides, it is still unknown what the purpose and target population of eHealth is and why we would use it. That is scary. Then, you can imagine why eHealth has low priority.* [Interview, manager primary health care center, March 9, 2016]

Entrepreneurs

Envisaged Role in eLabEL

The entrepreneurs' role in eLabEL was to bring in mature and evidence-based eHealth applications, in conjunction with patients and health care professionals, and to combine the different applications into one infrastructure via a single sign-on. To realize a sustainable, intelligent, and interoperable information and communication technology (ICT) infrastructure, which was necessary for eLabELs' mission, the individual applications as well as the infrastructure should meet the national and European requirements for data exchange, data safety, and data privacy. The entrepreneurs were also asked for knowledge and financial investments.

Small- and Medium-Sized Enterprises Incentive to Collaborate

Participating entrepreneurs started in eLabEL with the expectation that collaboration with research institutes would help them to enter a new market, that is, primary health care. They felt they needed to collaborate with other entrepreneurs as they were not big enough to enter the market on their own. They expected collaboration with research institutes as an important surplus value: it would add a scientific basis for their applications and therefore could create additional market value. They valued the intensive collaboration among the SMEs resulting in small alliances of 2 or 3 SMEs, as well as the experiences of participating in the project as a whole.

Small- and Medium-Sized Enterprises Need a Positive Business Case

During the project, it came to the fore that entrepreneurs did not have the technical knowledge that was needed to set up a sustainable interoperable ICT infrastructure and that their eHealth applications were neither fully mature nor evidence based. Furthermore, the entrepreneurs were continuously considering whether investments in eLabEL would result in future revenues (mainly in the short term). As the SMEs differed in their motivation and in weighing investments, it proved hard to create a shared vision statement on the integration of the different eHealth applications, the investment strategy, and a joined entity to assign intellectual properties to. The main reason for the struggles experienced in the cooperation among SMEs was that the SMEs differed in their convictions of future revenues because of uncertainties in the financial market and that it was not possible to make a positive business case. An individual investment in the eHealth infrastructure was considered as unwise and too risky by each SME, and therefore, they opted for a joint investment. However, the business case and corresponding business model should still be positive. Questions like who will pay, who is the customer, and who is the user were difficult to answer during the whole project, as the primary health care market was a new and therefore relatively unknown market for the SMEs. This resulted in continuous discussions on the business model. One of the SME's explained it as:

> *The health care market is unknown. Who should pay for it? How can we sell it? The Business Model is unclear. For medical care the health care insurance should pay. For non-medical care a patient or health care organization should pay. This is difficult in primary health care.* [Meeting entrepreneurs, September 16, 2014]

Health Care Insurers and Policy Makers

Envisaged Role in eLabEL

In the eLabEL project, the expected role of policy makers and health care insurers was that of enabling the health care professionals in experimenting with the use of eHealth in primary care. More precisely, we expected that health care insurers would provide financial support for the appointment of practice nurses.

Shared Ambition

Several discussions were held with health care insurers and policy makers. Time after time it was clear that we had a shared ambition: health care insurers and policy makers agreed that it was necessary to work on an integrated infrastructure for eHealth applications to transform traditional primary health care into technology-supported health care. In their view, the Dutch financial legislation offers prospects for financing eHealth applications as there are policy rules, conditions for reimbursement, and incentives for innovation.

Cost-Effectiveness Studies Are Needed

Simultaneously, health care insurers were reserved. They needed a business case and insight into cost-effectiveness of the infrastructure that we were developing before they would think about reimbursement or investments. We could not achieve this in the project and therefore, they did not want to support the project. It seemed that health care insurers were mostly interested in short-term effects. In actual practice, the Dutch regulations and legislation seemed to act inconsistently: they argue to stimulate eHealth on one side but require cost-effectiveness studies first on the other side. However, to carry out cost-effectiveness studies in real practice, implementation of eHealth needs legislation and financial regulations first.

Instead of investments by health care insurance, health care organizations themselves might be able to invest in eHealth-applications. However, in the Dutch pay-for-performance-based health care system, the use of eHealth applications that lower the number of consults will also lower the health care professionals' revenues. Actually, investments by the health care organizations will lead to lower costs for the health care insurers but also lower income for the health care practices. This was explained by one of the managers as follows:

> *eLabEL is aimed at more efficient care and better quality of care, with the ideal result that patients are more satisfied. But, this should not result in cost savings only for the health care insurance sector...The shared savings principle might be worthwhile.* [Meeting entrepreneurs and primary health care managers, November 6, 2014]

Discussion

About 4 years after starting the eLabEL project, we conclude that, despite the hard work and collaboration of many stakeholders, it was not possible to implement eHealth in these Living Labs at this moment in time. One might say that the eLabEL consortium failed in its ambition. However, we gathered in-depth information about the complexity of innovations in primary health care that can help many researchers, entrepreneurs, and policy makers in setting up the next initiatives on this topic. Our experiences in eLabEL taught us that successful use of eHealth needs more than enthusiastic partners. Successful use also depends on the efforts of all stakeholders, their willingness to invest time or money, and shared vision statements. Although it is not easy because of different contexts, values, and expectations, based on the experiences inside and outside eLabEL, we still believe that collaboration between all 4 groups of stakeholders, that is, patients, health care professionals, entrepreneurs, and health care insurers or policy makers is essential. Moreover, we argue that policy, especially the health care insurance market, should be added as a field to Eysenbach's definition of eHealth [9].

Were we naive when starting eLabEL? We might be: we knew we were ambitious, but looking back, we realize we had unrealistic expectations and our goals were not specific enough. Nevertheless, SMEs were willing to collaborate and to invest as they were ready to step into a new market. Additionally, health care organizations also felt the urge to participate. What we did not foresee was the struggle (1) to convince health care insurers to support health care professionals in our project and (2) to create a positive business model. In fact, it was those factors that led to an impasse: without commitment of health care centers or insurers, no positive business model could be created, and SMEs could not invest in the eLabEL infrastructure. However, without investments in the eLabEL infrastructure, health care professionals were not convinced of its added value. Moreover, without financial support by health care insurers, they were not motivated or able to use it. This made it impossible for the researchers to collect the evidence that health care insurers were asking for.

With this paper, we aimed to show the process of adoption of eHealth tools in primary care illustrated by the eLabEL project, which provided us a rich qualitative dataset. However, our study has some limitations. The main limitation is the involvement of patients, health care insurers, and policy makers. Patients were supposed to be part of the Living Lab. They were actively involved at the start of the project. Slow progress in development and implementation changed the focus of the project and resulted in less involvement of patients. Health care insurers and policy makers were not part of the Living Lab and were therefore less involved in the project. Nevertheless, it is clear from our observations that a number of actions should be done differently in future projects to enforce the implementation of eHealth in primary care. These actions will be discussed in the following sections and are listed in Textbox 2.

It turned out that it is not easy to actively involve patients in research projects with an eHealth-topic. Considering the importance of their participation, especially in eHealth projects [20], they should be supported in participating in the project. Wildevuur et al (2017) recently published 4 preconditions for enhancing the partnership in ICT-enabled person-centered care [21]. In addition, incentives for care practices seem to be needed to recruit patients for participation. Patient's expectations of the benefits of using eHealth play an essential role in their actual use. Therefore, it is essential that Living Labs as set up in eLabEL awake patients' interest by offering relevant eHealth tools. Patients' expectations are not only dependent on the technology but also on the way in which general practices offer, promote, and use it [22,23]. Care professionals should be supported in informing patients about the possibilities, uses, and reasons for implementation, focusing on the benefits eHealth can bring. However, whether patients will actually use eHealth will always be personal and differ among patients. Monitoring which patients benefit the most from the use of eHealth and those who do not, seems to be important to develop optimal implementation strategies.

From the care organizations' perspective, it appeared that health care managers and professionals were not ready to implement eHealth tools without support. Implementing eHealth requires it to be a fundamental part of the mission and vision of the health care organization. Only then decisions on budget and support can be made. We found that the process of adapting and implementing eHealth is too complicated to organize next to regular care-giving activities for primary health care professionals. In addition, we learned that involving a primary health care center in the plans is not sufficient.

Textbox 2. Lessons learned from a Living Lab on broad adoption of electronic health (eHealth) in primary health care.

- Patients need support to actively participate in eHealth projects, and those projects need to be relevant for the patients
- Incentives for care practices are needed to recruit patients for participation in eHealth projects
- Primary care practices need support to adequately inform patients and monitor which patients benefit from the use of eHealth
- The community in which a primary health care system operates needs to be involved in eHealth projects
- Primary care practices need support and managerial power for the implementation and innovation processes
- Collaborated eHealth entrepreneurs need trust in each other, shared vision statements, and early commitment to short- and long-term goals
- A business model concept is needed early in eHealth projects and essential for collaboration
- Strategies are needed focusing on financial models that stimulate innovation and on requirements needed for societal innovations
- Patients, primary health care professionals, entrepreneurs, and government need to act together in eHealth projects

The use of eHealth goes beyond the own practice borders as primary health care professionals often operate close to other health care professionals in their region. This makes the innovations even more complex as those parties also need to be involved [24]. Nowadays, it is common knowledge that implementation of innovations, including eHealth, is difficult and progresses only slowly [25-27]. Lau et al (2016) stated in a recent systematic review of reviews that *implementing any type of change in primary care is likely to be complex* and that *relevant barriers and facilitators are dynamic and likely to change over time* [25]. Theoretical models show that the innovation process or implementation infrastructure are important parts of implementation, next to the intervention characteristics, the organizational structure, the context, and the individuals [8,28]. Moreover, it is shown that a greater knowledge of essential adjustments in health care provider workflow, roles, and responsibilities is needed [29]. Our study provides a real-world view on these topics showing that Dutch primary care organizations, mainly small organizations, do not have the managerial power that is needed for complex innovations such as large-scale eHealth implementation. Primary care organizations probably will benefit from infrastructure that support them in the implementation process.

From the entrepreneurs' perspective, it is important to have shared vision statements and a business model concept as early as possible. Clearly defined short- and long-term goals are needed. In addition, trust in each other and commitment of all parties is important. Knowledge of the potential cost-effectiveness of eHealth is an important requirement for all stakeholders. The use of an early health technology assessment can provide insight into potential outcomes, drivers, and barriers. Moreover, we should realize that SMEs might have difficulties, due to lack of knowledge, in developing interoperable eHealth, when facing different Dutch and European requirements on data safety, data exchange, and data privacy. Furthermore, in developing, adapting, selecting, and implementing eHealth tools, they should work as closely as possible with the end user, that is, health care professionals and patients. Active user involvement is a time-consuming process. Developers should balance the need for input from users, with the availability of resources such as time and funding [30]. To remain competitive within a fast-moving market, it is important to develop quickly [31]. However, we recommend that the need assessment phase should not be neglected; this seems of major importance for the development of eHealth from which patients can experience benefits and might be an important trigger to actually use eHealth.

The implementation of eHealth is not yet a fully recognized aspect of primary health care organizations, which makes it difficult to fit eHealth locally. Furthermore, inconsistencies in policy rules hinder improvements and innovations. For projects such as eLabEL, it would help when policy makers and health care insurers would allow experiments in which standard regulations can be (partly) neglected to fully explore new financing models. This can only be arranged when policy makers and health care insurers are involved from the beginning of the project. However, such experiments are not a structural solution for broad-scale implementation of eHealth. Moreover, financial support does not guarantee the large-scale use of eHealth [32]. Broad-scale implementation will need strategies that not only focus on financial models that stimulate innovation but also on requirements needed for societal innovations [33,34].

In conclusion, we believe that for optimal and sustainable use of eHealth, patients should be actively involved, primary health care professionals need to be encouraged in their management, entrepreneurs should work closely with health care professionals and patients, and government needs to focus on new health care models stimulating innovations. Only when all these parties act together, starting in local communities with a small range of eHealth tools, the potential of eHealth will be realized.

Conflicts of Interest

LV works at one of the SMEs that participated in eLabEL. Other authors have no conflicts of interest to declare.

Multimedia Appendix 1

Theoretical models.

[PDF File (Adobe PDF File), 14KB - jmir_v20i3e83_app1.pdf]

References

1. Henderson C, Knapp M, Fernández JL, Beecham J, Hirani S, Cartwright M, Whole System Demonstrator evaluation team. Cost effectiveness of telehealth for patients with long term conditions (Whole Systems Demonstrator telehealth questionnaire study): nested economic evaluation in a pragmatic, cluster randomised controlled trial. Br Med J 2013 Mar 20;346:f1035 [FREE Full text] [Medline: 23520339]
2. Flodgren G, Rachas A, Farmer A, Inzitari M, Shepperd S. Interactive telemedicine: effects on professional practice and health care outcomes. Cochrane Database Syst Rev 2015 Sep 07(9):CD002098. [doi: 10.1002/14651858.CD002098.pub2] [Medline: 26343551]
3. de Lusignan S, Mold F, Sheikh A, Majeed A, Wyatt JC, Quinn T, et al. Patients' online access to their electronic health records and linked online services: a systematic interpretative review. BMJ Open 2014 Sep 08;4(9):e006021 [FREE Full text] [doi: 10.1136/bmjopen-2014-006021] [Medline: 25200561]

4. Wildevuur SE, Simonse LW. Information and communication technology-enabled person-centered care for the "big five" chronic conditions: scoping review. J Med Internet Res 2015 Mar 27;17(3):e77 [FREE Full text] [doi: 10.2196/jmir.3687] [Medline: 25831199]
5. Peeters J, Wiegers T, Friele R. How technology in care at home affects patient self-care and self-management: a scoping review. Int J Environ Res Public Health 2013 Oct 29;10(11):5541-5564 [FREE Full text] [doi: 10.3390/ijerph10115541] [Medline: 24173139]
6. van Dyk L. A review of telehealth service implementation frameworks. Int J Environ Res Public Health 2014 Jan 23;11(2):1279-1298 [FREE Full text] [doi: 10.3390/ijerph110201279] [Medline: 24464237]
7. 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]
8. Damschroder L, Aron D, Keith R, Kirsh S, Alexander J, Lowery J. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
9. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
10. Vitacca M, Mazzù M, Scalvini S. Sociotechnical and organizational challenges to wider e-Health implementation. Chron Respir Dis 2009;6(2):91-97. [Medline: 19411570]
11. Van Velsen L, Wentzel J, Van Gemert-Pijnen JE. Designing eHealth that matters via a multidisciplinary requirements development approach. JMIR Res Protoc 2013 Jun 24;2(1):e21 [FREE Full text] [doi: 10.2196/resprot.2547] [Medline: 23796508]
12. Vermeulen J, Huygens M, de Witte L, Oude Nijeweme-d'Hollosy W, van Velsen L, Swinkels I, et al. eLabEL: Technology-supported Living Labs in Primary Care. In: eTELEMED 2015: The Seventh International Conference on eHealth, Telemedicine, and Social Medicine. eTELEMED 2015: IARIA; 2015 Presented at: eTelemed; 22-27 february; Lisbon Portugal.
13. ENoLL. Open Living Labs. What is a Living Lab? URL: http://www.openlivinglabs.eu/FAQ [accessed 2017-09-13] [WebCite Cache ID 6tRaSbKGj]
14. Huygens M, Vermeulen J, Swinkels I, Friele R, van Schayk O, de Witte L. Expectations and needs of patients with a chronic disease toward self-management and eHealth for self-management purposes. BMC Health Serv Res 2016 Dec 08;16:232 [FREE Full text] [doi: 10.1186/s12913-016-1484-5] [Medline: 27391471]
15. Nijeweme-d'Hollosy WO, van Velsen L, Huygens M, Hermens H. Requirements for and barriers towards interoperable eHealth technology in primary care. IEEE Internet Comput 2015 Jul;19(4):10-19. [doi: 10.1109/MIC.2015.53]
16. Wagemakers A, Koelen M, Lezwijn J, Vaandrager L. Coordinated action checklist: a tool for partnerships to facilitate and evaluate community health promotion. Glob Health Promot 2010 Sep;17(3):17-28. [doi: 10.1177/1757975910375166] [Medline: 21495437]
17. Nyström A, Leminen S, Westerlund M, Kortelainen M. Actor roles and role patterns influencing innovation in living labs. Ind Market Manag 2014 Apr;43(3):483-495. [doi: 10.1016/j.indmarman.2013.12.016]
18. Geels F. Technological transitions as evolutionary reconfiguration processes: a multi-level perspective and a case-study. Res Policy 2002 Dec;31(8-9):1257-1274. [doi: 10.1016/S0048-7333(02)00062-8]
19. 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]
20. Latulippe K, Hamel C, Giroux D. Social health inequalities and eHealth: a literature review with qualitative synthesis of theoretical and empirical studies. J Med Internet Res 2017 Apr 27;19(4):e136 [FREE Full text] [doi: 10.2196/jmir.6731] [Medline: 28450271]
21. Wildevuur S, Thomese F, Ferguson J, Klink A. Information and communication technologies to support chronic disease self-management: preconditions for enhancing the partnership in person-centered care. J Participat Med 2017 Jul 07;9(1):e14. [doi: 10.2196/jopm.8846]
22. Wells S, Rozenblum R, Park A, Dunn M, Bates D. Organizational strategies for promoting patient and provider uptake of personal health records. J Am Med Inform Assoc 2015 Jan;22(1):213-222 [FREE Full text] [doi: 10.1136/amiajnl-2014-003055] [Medline: 25326601]
23. Wolcott V, Agarwal R, Nelson D. Is provider-secure messaging associated with patient messaging behavior? J Med Internet Res 2017 Apr 06;19(4):e103 [FREE Full text] [doi: 10.2196/jmir.6804] [Medline: 28385681]
24. Timmermans S, Berg M. Standardization in action: achieving local universality through medical protocols. Soc Stud Sci 2016 Jun 29;27(2):273-305. [doi: 10.1177/030631297027002003]
25. Lau R, Stevenson F, Ong BN, Dziedzic K, Treweek S, Eldridge S, et al. Achieving change in primary care--causes of the evidence to practice gap: systematic reviews of reviews. Implement Sci 2016 Mar 22;11:40 [FREE Full text] [doi: 10.1186/s13012-016-0396-4] [Medline: 27001107]

26. Peeters J, Krijgsman J, Brabers A, De Jong JD, Friele R. Use and uptake of eHealth in General Practice: a cross-sectional survey and focus group study among health care users and general practitioners. JMIR Med Inform 2016 Apr 06;4(2):e11 [FREE Full text] [doi: 10.2196/medinform.4515] [Medline: 27052805]
27. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Dec 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
28. Feldstein A, Glasgow R. A practical, robust implementation and sustainability model (PRISM) for integrating research findings into practice. Jt Comm J Qual Patient Saf 2008 Apr;34(4):228-243. [Medline: 18468362]
29. Irizarry T, DeVito DA, Curran C. Patient portals and patient engagement: a state of the science review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
30. de Beurs D, van Bruinessen I, Noordman J, Friele R, van Dulmen S. Active involvement of end users when developing web-based mental health interventions. Front Psychiatry 2017;8:72 [FREE Full text] [doi: 10.3389/fpsyt.2017.00072] [Medline: 28515699]
31. Money A, Barnett J, Kuljis J, Craven M, Martin J, Young T. The role of the user within the medical device design and development process: medical device manufacturers' perspectives. BMC Med Inform Decis Mak 2011 Feb 28;11:15 [FREE Full text] [doi: 10.1186/1472-6947-11-15] [Medline: 21356097]
32. Huygens MWJ, Swinkels ICS, Verheij RA, Friele RD, van Schayck OCP, de Witte LP. Understanding the use of email consultation in primary care using a retrospective observational study with data of Dutch electronic health records. BMJ Open 2018 Jan 21;8(1):e019233. [doi: 10.1136/bmjopen-2017-019233] [Medline: 29358442]
33. Janssen M. Situated novelty: a study on healthcare innovation and its governance: Health Care Governance (HCG) PhD dissertation. Rotterdam The Netherlands: Erasmus University; 2016.
34. Krijgsman J, Swinkels I, van Lettow B, de Jong J, Out K, Friele R, et al. eHealth-monitor 2016. Den Haag, Utrecht: Nictiz/ Nivel; 2016.

Abbreviations

COPD: chronic obstructive pulmonary disease
eHealth: electronic health
ICT: information and communication technology
SMEs: small- and medium-sized enterprises

Edited by G Eysenbach; submitted 04.10.17; peer-reviewed by J Ross, N Bragazzi, S Wildevuur; comments to author 16.11.17; revised version received 22.12.17; accepted 02.01.18; published 29.03.18.

Please cite as:
Swinkels ICS, Huygens MWJ, Schoenmakers TM, Oude Nijeweme-D'Hollosy W, van Velsen L, Vermeulen J, Schoone-Harmsen M, Jansen YJFM, van Schayck OCP, Friele R, de Witte L
Lessons Learned From a Living Lab on the Broad Adoption of eHealth in Primary Health Care
J Med Internet Res 2018;20(3):e83
URL: http://www.jmir.org/2018/3/e83/
doi:10.2196/jmir.9110
PMID:29599108

Original Paper

Optimizing Tailored Communications for Health Risk Assessment: A Randomized Factorial Experiment of the Effects of Expectancy Priming, Autonomy Support, and Exemplification

Carmina G Valle[1,2], MPH, PhD; Tara L Queen[2*], PhD; Barbara A Martin[2*], MPH; Kurt M Ribisl[2,3], PhD; Deborah K Mayer[2,4], RN, PhD, AOCN, FAAN; Deborah F Tate[1,2,3], PhD

[1]Department of Nutrition, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States
[2]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States
[3]Department of Health Behavior, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States
[4]School of Nursing, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States
[*]these authors contributed equally

Corresponding Author:
Carmina G Valle, MPH, PhD
Department of Nutrition
Gillings School of Global Public Health
University of North Carolina at Chapel Hill
135 Dauer Drive, MHRC 2004
Campus Box #7461
Chapel Hill, NC, 27599 7461
United States
Phone: 1 919 843 6886
Fax: 1 919 843 6663
Email: carmina.valle@unc.edu

Abstract

Background: Health risk assessments with tailored feedback plus health education have been shown to be effective for promoting health behavior change. However, there is limited evidence to guide the development and delivery of online automated tailored feedback.

Objective: The goal of this study was to optimize tailored feedback messages for an online health risk assessment to promote enhanced user engagement, self-efficacy, and behavioral intentions for engaging in healthy behaviors. We examined the effects of three theory-based message factors used in developing tailored feedback messages on levels of engagement, self-efficacy, and behavioral intentions.

Methods: We conducted a randomized factorial experiment to test three different components of tailored feedback messages: tailored expectancy priming, autonomy support, and use of an exemplar. Individuals (N=1945) were recruited via Amazon Mechanical Turk and randomly assigned to one of eight different experimental conditions within one of four behavioral assessment and feedback modules (tobacco use, physical activity [PA], eating habits, and weight). Participants reported self-efficacy and behavioral intentions pre- and postcompletion of an online health behavior assessment with tailored feedback. Engagement and message perceptions were assessed at follow-up.

Results: For the tobacco module, there was a significant main effect of the exemplar factor (P=.04); participants who received exemplar messages (mean 3.31, SE 0.060) rated their self-efficacy to quit tobacco higher than those who did not receive exemplar messages (mean 3.14, SE 0.057). There was a three-way interaction between the effect of message conditions on self-efficacy to quit tobacco (P=.02), such that messages with tailored priming and an exemplar had the greatest impact on self-efficacy to quit tobacco. Across PA, eating habits, and weight modules, there was a three-way interaction among conditions on self-efficacy (P=.048). The highest self-efficacy scores were reported among those who were in the standard priming condition and received both autonomy supportive and exemplar messages. In the PA module, autonomy supportive messages had a stronger effect on self-efficacy for PA in the standard priming condition. For PA, eating habits, and weight-related behaviors, the main effect of exemplar messages on behavioral intentions was in the hypothesized direction but did not reach statistical significance (P=.08).

When comparing the main effects of different message conditions, there were no differences in engagement and message perceptions.

Conclusions: Findings suggest that tailored feedback messages that use exemplars helped improve self-efficacy related to tobacco cessation, PA, eating habits, and weight control. Combining standard priming and autonomy supportive message components shows potential for optimizing tailored feedback for tobacco cessation and PA behaviors.

(J Med Internet Res 2018;20(3):e63) doi:10.2196/jmir.7613

KEYWORDS

health communication; feedback; eHealth; health risk assessment; health behavior; intention; self-efficacy; personal autonomy

Introduction

Background

Health risk assessments plus feedback and additional educational approaches have been shown to be effective for supporting health behavior change [1]. Although many early health risk assessments were delivered via print [2], increasingly, these assessments are delivered via the Web. Web-based tailored interventions have demonstrated efficacy in promoting healthy behavior changes [3,4]. Online delivery of health risk assessments and tailored feedback present opportunities for greater reach and dissemination of effective health interventions that have the potential to lower costs, alleviate barriers to participation, and facilitate adherence to healthy behaviors. Although a large body of evidence supports the efficacy of tailored communications for promoting healthy behaviors (eg, [2,4-7]), there is limited evidence to guide the *optimal presentation of automated tailored feedback* on health behavior data to individuals [8]. Indeed, systematic reviews have called for researchers to provide enhanced descriptions of tailoring criteria and message design to optimize the use of Web-based tailored interventions to promote behavior change [4,9].

The Carolina Health Assessment and Resource Tool (CHART) is an online health behavior risk assessment tool developed at the University of North Carolina that includes various assessments of health behaviors related to chronic diseases that comprise the leading causes of premature death in the United States, along with evidence-based, theory-guided tailored feedback message libraries [2,4,10]. Individuals complete a baseline assessment on their current status in meeting the national recommendation for a specific health behavior (eg, physical activity [PA] and being tobacco-free), as well as other theory- and evidence-based psychosocial factors related to the health behaviors. Responses to this assessment are used within CHART to create a tailored feedback report (*personalized report*) based on an individuals' reported current behavior, readiness to change, perceived barriers, and social support [10].

Message content and delivery format of tailored feedback reports, such as those offered by CHART, are critical components of online health messages that may affect an individual's evaluation of a message, website use, and subsequent behavior change. To date, tailored health interventions have commonly used the major health behavior theories to guide message content (eg, social cognitive theory [SCT], transtheoretical model, and health belief model) [2,11,12]. Meta-analyses of tailored health behavior change interventions indicate that tailoring on more theoretical constructs (ie, 4-5 or more) in addition to behavior and demographics may improve the effectiveness of tailored interventions [2]. However, these theories contain many individual constructs, and there is a need to identify the specific message components of these tailored interventions in a systematic manner that will enhance understanding of the most effective message features and guide optimization and future testing of this set of features [13].

A growing literature has encouraged the use of the multiphase optimization strategy (MOST) framework to elucidate the active ingredients of interventions [13-15]. Thus, this study was designed to examine multiple message components (factors) of CHART personalized reports. The goals were to optimize the existing tailored feedback to promote self-efficacy and behavioral intentions for engaging in healthy behaviors and to enhance user engagement on the dimension of subjective experience. Guided by a recent systematic review and conceptual framework on engagement with digital behavior change interventions [16], our focus was on the experiential aspects of engagement, characterized by interest, affect, and attention (eg, self-report measures of perceptions of effectiveness, information quality, and attractiveness), and how engagement with the tailored feedback might be impacted by the content and delivery of the tailored feedback. Given that a single administration of a health risk assessment plus feedback, without additional intervention approaches, was unlikely to affect behavior change, we focused on self-efficacy and behavioral intentions. Both of these psychosocial constructs are key components of health behavior theories (eg, SCT and theory of planned behavior) and have been shown to be proximal determinants and predictors of behavior change [17,18]. Consistent with the screening phase of the MOST framework, we used a factorial design to allow for testing of the main effects on outcomes, as well as prespecified interactions. Our focus was on three specific message factors: expectancy priming, autonomy support, and exemplification, which are detailed below.

Expectancy Priming

Individuals may vary on their tailoring-related expectancies, or the value or benefit that one may expect from tailored communications, and these expectancies are changeable [19]. Webb et al [20] demonstrated that individuals' baseline expectancies about tailoring moderated the effect of personalized smoking cessation booklets on readiness to quit smoking, such that extensively personalized materials produced greater effects on readiness to quit among those with more positive expectancies about tailoring. A follow-up study showed that

expectancy priming to manipulate tailoring-related expectancies (ie, making it clear that a message is either a standard one or personally tailored for that individual) can enhance the value or benefit that participants expect from either standard or tailored materials, improve ratings of message content, enhance readiness to change, and promote behavior change [19]. Thus, expectancy priming may influence the effect of tailored interventions such as feedback reports from health risk assessments. For this study, before receiving their tailored feedback report, participants were randomized to receive either a priming message explicitly stating that their feedback report was personally tailored for them, or they received a priming message stating that the feedback included a standard report.

Autonomy Support

Behavioral interventions and tailored messages using self-determination theory (SDT) [21] as a guiding framework have demonstrated effectiveness for improving health behaviors such as PA [22], fruit and vegetable intake [23], and weight control [24]. SDT distinguishes between autonomous motivation and controlled motivation and posits that the *type of motivation*, rather than amount, is more influential on behavior [21,25]. Behaviors are autonomously motivated when actions result from conscious choice and are personally relevant, whereas controlled motivation involves engaging in a behavior because of perceived external pressures [26]. When individuals are autonomously motivated, behavior changes have been shown to be more effective and sustained [27]. Many health behavior interventions have focused on increasing autonomous motivation and in turn improved behavioral outcomes [22,28,29]. Often these interventions have encouraged autonomous motivation by incorporating autonomy supportive behavioral strategies based on SDT, including providing several options for change, supporting a sense of choice, eliciting an individual's emotions, providing rationale for the importance of a behavior, and exploring the relevance of behaviors for an individual's values and goals [30]. Thus, participants in this study were randomized to receive tailored feedback messages that were either autonomy supportive or used more directive language (ie, existing CHART feedback). On the basis of recommendations for enhancing autonomous motivation [26], the autonomy supportive messages offered a sense of choice or menu of options for change and encouraged participants to consider their own motivations and solutions to barriers, whereas the directive messages more explicitly told participants what to do.

Exemplification

Self-efficacy, or confidence in one's ability to take action or perform a behavior in the face of obstacles [31], is one of the most commonly targeted theoretical determinants in behavioral interventions and is central to multiple theories of behavior change (eg, SCT, transtheoretical model, and health belief model). Vicarious experiences and verbal persuasion are specific strategies relevant to health communication [32] and similar to exemplification used in media [33], which can be used to promote self-efficacy. Messages presented with an exemplar (eg, role model for the behavior) enable individuals to observe others perform an activity successfully, learn from vicarious experience, communicate positive outcomes of a behavior, and can encourage self-efficacy and in turn lead to behavior change [34-36]. Previous studies have shown that messages featuring exemplars, or role models, improve perceptions of self-efficacy and have the potential to enhance attractiveness of a message and promote positive behavior change [37]. In this study, participants were randomized to receive messages that featured an exemplar or did not use an exemplar.

Using manipulations of these three different message characteristics (expectancy priming, autonomy support, and exemplification), this study examined whether tailoring feedback messages using three different approaches improved self-efficacy and behavioral intentions to adhere to recommended health behaviors. As previous research has indicated that engagement with or subjective perceptions of tailored messages (eg, perceived message relevance) may mediate or explain the mechanism of tailoring effects [38-40], we also examined the effects of message factors on engagement, or subjective experiences related to attention, interest, and affect with tailored feedback messages. The overall goal was to inform the selection of the most effective messages for use in future CHART personalized reports and to expand the scientific basis for the optimal presentation of tailored feedback. We hypothesized that tailored expectancy priming, autonomy supportive messages, and messages with exemplars would be more effective for improving self-efficacy, behavioral intentions, and engagement compared with messages without these features.

Methods

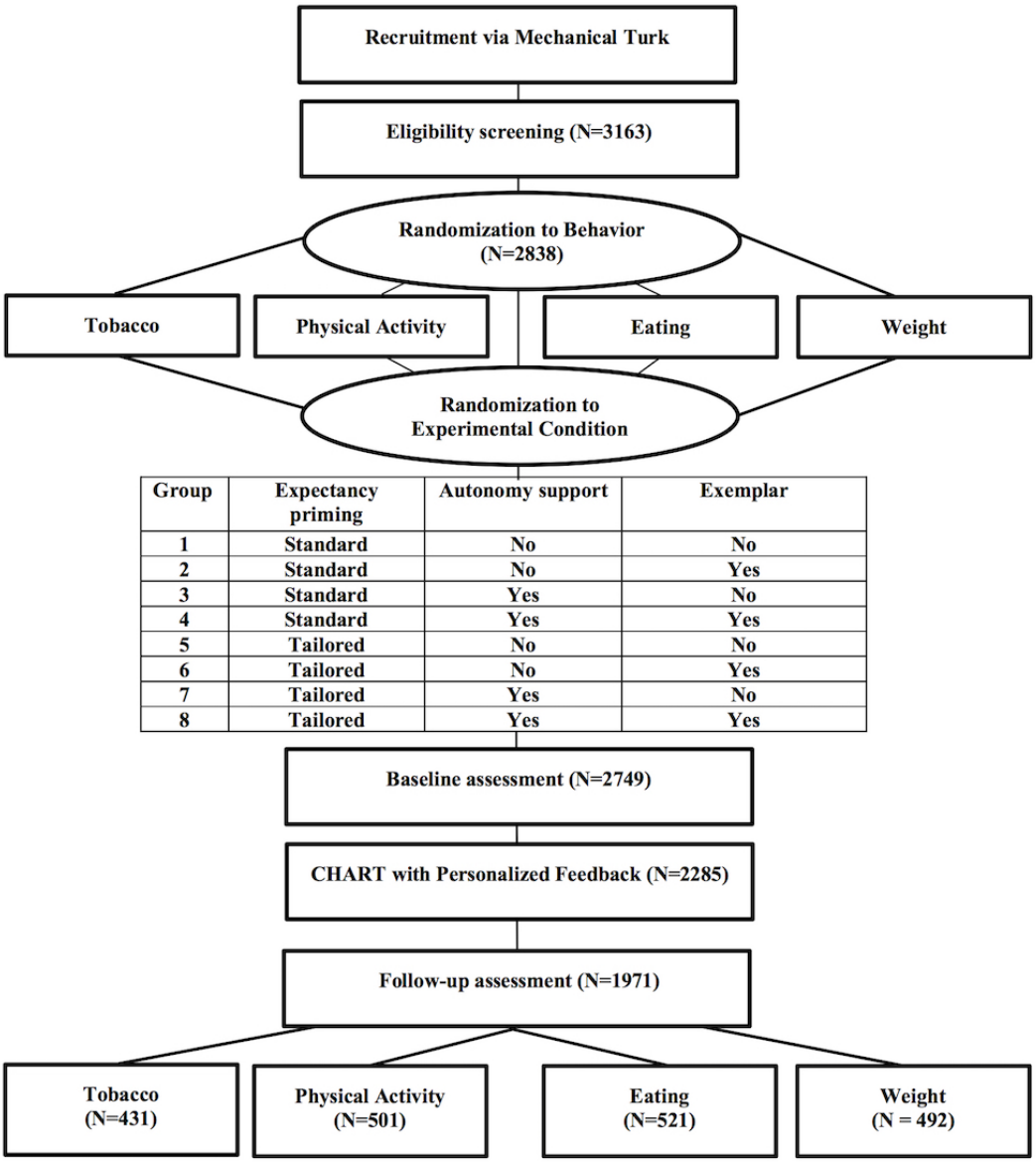
Participants and Recruitment

Participants were recruited from January 2016 to February 2016 through Amazon Mechanical Turk (MTurk), a website where tasks are crowdsourced to employees, called workers, who receive compensation for completing Human Intelligence Tasks (HITs) [41]. MTurk has been used in a number of different research studies to collect a diversity of information such as health knowledge of ovarian cancer [42], ways to increase PA [43], and to measure body image [44] and the perceived harmfulness of tobacco products [45,46]. It is a useful tool for behavioral researchers because of its low cost, diverse worker population, and speed of data collection [41]. In MTurk, this study advertised the HIT as an opportunity to share opinions about a health survey and described the task as needing feedback about an online health website. Participants met the following initial eligibility criteria: aged 18 years or older, had Internet access, reside in the United States, able to communicate in English, and HIT approval rate (ie, percentage of worker's completed HITs that have been approved by requestors) greater than or equal to 90%. Individuals were given US $1.25 for successful completion of the assignment, as detailed below. This study was reviewed and exempted by the institutional review board of the University of North Carolina at Chapel Hill (IRB #14-2873).

Procedure and Design

We conducted pre- posttest experiments to test the effects of eight different message conditions on self-efficacy and intentions to engage in four different health behaviors.

Figure 1. Flow of study participants in randomized 2x2x2 factorial experiment. CHART: Carolina Health Assessment and Resource Tool.



This study used a factorial design testing three experimental factors, with each factor having two levels of message characteristics (2x2x2): (1) expectancy priming before feedback delivery (standard vs tailored), (2) autonomy supportive messages (presence vs absence), and (3) use of exemplars (presence vs absence; Figure 1). We tested tailored feedback messages for four behavioral modules in CHART (tobacco use, PA, eating habits, and weight). For each of these four behaviors, the message characteristics were fully crossed (3 factors by 2 levels). Thus, for the independent variables of message condition, eight different cells were generated. Participants were randomized to one of the four behaviors (or 1 of 3, or 1 of 2, depending on eligibility, or assigned to 1 if only eligible for 1) and then subsequently randomly assigned, with equal probability, to one of eight experimental groups (Figure 1). The procedures from initial recruitment in MTurk through study completion are outlined below. In MTurk, the HIT indicated eligibility requirements to participate and provided a URL to begin the HIT. Participants were then directed through a series of online questionnaires and websites in the following order:

- This first questionnaire was an online screener with questions that asked individuals to report their gender and current smoking habits, PA behaviors, intake of fruits and vegetables, height, and weight. This screener identified participants that met additional eligibility criteria of not meeting national recommendations for at least one of four cancer prevention–related health behaviors: current tobacco use (smoker), PA (ie, less than 150 min of moderate exercise a week), fruit and vegetable consumption (ie, consuming less than 2 servings of fruit or 3 servings of vegetables daily), and weight status (ie, overweight or obese, body mass index ≥25).
- On the basis of eligibility, participants were randomized into one of the up to four health behaviors for which they were not meeting national recommendations. For instance, individuals who were not meeting national

recommendations for all of the four behaviors were randomly assigned to one of the four modules. Individuals who were not meeting national recommendations for two behaviors (eg, PA and eating habits) were randomly assigned to one of these two modules. Respondents who did not meet the national recommendation for only one behavior were assigned to the corresponding module. Within each behavior, individuals were randomly assigned to one of eight different message conditions (Figure 1) based on three factors: expectancy priming (standard or tailored), autonomy support (yes or no), and exemplar (yes or no).

- Upon randomization, participants completed an online baseline questionnaire related to the specific behavioral module, which assessed self-efficacy and behavioral intentions. Participants received a unique user password after completing the questionnaire.
- Next, participants were directed to the CHART website [47], where they entered their password and completed an assessment questionnaire related to their assigned health behavior and a demographics questionnaire. Questions included standard items used in the CHART assessments (ie, current health behavior, readiness to change, barriers to engaging in the specific health behavior, and social support).
- Upon completion of the CHART assessment, participants received a personalized report that was tailored-based on preexisting tailoring variables programmed in CHART (current health behavior, readiness to change, barriers, and social support) and included messages with features consistent with one of the eight randomized conditions.
- Participants were instructed to read through the report and to click on a link that directed them to a password-protected final online questionnaire that asked about their opinions on the personalized report.
- Once participants completed the final questionnaire, they received a HIT completion code. After entering the completion code on the MTurk website, participants received US $1.25 for their time. Mean study completion time was 15.6 min (standard deviation [SD] 9.9).

Data collection for this study was completed in four cohorts.

Experimental Conditions

Multimedia Appendix 1 provides examples of the message text in the various experimental conditions delivered through the personalized reports, which are detailed below.

Expectancy Priming

In the introduction to the personalized report, participants were randomized to receive either a standard priming or tailored priming message that described the contents of the report. The landing page used common graphics and language to direct individuals to click on a link to access their personalized report. The description of the personalized report differed between conditions and was adapted from previous research [20] (Multimedia Appendix 1). The standard priming condition indicated that the report was based on research that may help people meet recommendations. Language in the tailored priming condition stressed that the report was tailored especially for the individual and designed to meet their unique needs.

Autonomy Support

Participants were randomized to receive tailored feedback that consisted of either messages designed to be autonomy supportive or messages without a focus on autonomy support (standard existing CHART messages or exemplar messages). Tailored feedback in the autonomy supportive condition used language that was less directive and encouraged individuals to consider their own preferences and options. Original CHART-tailored feedback messages were revised and rephrased to ask more open-ended questions with the goal of promoting autonomy and choice (Multimedia Appendix 1). For instance, instead of directing individuals to “Fit weighing into your daily routine by stepping on the scale every morning when you get up,” the autonomy supportive messages asked, “How can you fit weighing into your daily routine? How about stepping on the scale every morning when you get up?”

Exemplar

In the exemplar condition, participants received messages that used gender-matched descriptions of a man (Bill) or woman (Rachel) who had a similar behavioral profile and had successfully made changes to meet the recommended behavioral goal. Tailored feedback in the original CHART message library was adapted to include Bill or Rachel as a role model for working toward improving their behaviors. For instance, an original CHART message read as follows: “Changing what you eat is not always easy. But, you can do it! Start with a goal you know you can reach. Small changes, like swapping sweetened drinks for water, can make a big difference to your health.” This same message was revised to include a role model for the behavior: “Changing what you eat is not always easy. But, you can do it! Like you, Rachel [Bill] had challenges that were getting in the way of her [his] healthy eating. She [He] started with a goal she [he] knew she [he] could reach and found that small changes, like swapping cookies for fruit, made a big difference to her [his] health.”

Autonomy Support x Exemplar

For participants that were randomized to this condition, tailored feedback combined messages that were autonomy supportive and included an exemplar (Multimedia Appendix 1). This was operationalized by using messages that were nondirective and asked questions that encouraged reflection, while also including Bill or Rachel as a role model exemplifying positive behaviors. For example, a standard tailored message regarding weight read as follows: “Try to fit veggies into every meal! Eating vegetables, especially those that are brightly colored, may help protect against heart disease and stroke.” This same message was revised to the following: “Have you thought about trying to fit veggies into every meal? Rachel [Bill] decided to eat more vegetables, especially brightly colored ones, since they can help protect her [him] against heart disease and stroke.”

Measures

Primary Outcomes

Self-efficacy was assessed at baseline and after receipt of the personalized report using a single item adapted from previous studies [48,49] that asked participants "How confident are you that you can..." (1) Quit smoking or stop using smokeless tobacco products?, (2) Get the recommended amount of PA each week?, (3) Eat at least (5 for women, 5½ for men) cups of fruit and vegetables each day?, or (4) Control your weight? Responses ranged from 1 (*not at all confident*) to 5 (*extremely confident*).

Behavioral intentions were measured at baseline and following receipt of the personalized report with two (tobacco, weight) or four items (eating habits, PA) on a 7-point scale from *1 (strongly disagree) to 7 (strongly agree).* Items were adapted from previous measures [50-53] and asked participants to indicate the extent to which they agree or disagree with statements about their intentions to engage in a health behavior goal over the next month (eg, I intend to exercise regularly over the next month. I will try to exercise regularly over the next month.). Measures for all modules used the stems "I intend to..." and "I will try to...," with the behavioral goals matching those appearing in the specific modules (ie, quit smoking and aim for a healthy weight). For the PA and eating habits measures, four items were used to assess behavioral intentions related to two behavioral goals (ie, exercise regularly, get at least 150 min of PA each week, eat at least two cups of fruits each day, and eat at least three cups of vegetables each day). Items were averaged for each behavior (Cronbach alpha=.92-.96).

Engagement and Perceptions of Personalized Report Messages

Perceived message relevance, which has been shown to be related to tailoring [6] and a mediator of behavior change [38,39], was measured with two items adapted from previous studies of tailored messages [38-40]. Participants were asked to rate how strongly they disagree or agree with the following statements: (1) "The information in the personalized report seemed to be written personally for me" and (2) "The information in the personalized report applied to my life." Responses were on a 5-point scale from 1 (*strongly disagree*) to 5 (*strongly agree*) and were averaged across the two items (Cronbach alpha=.80).

Perceived informativeness was assessed using a 2-item scale adapted from Cho and Boster [54] that asked participants to rate their agreement with statements on a 5-point scale. Statements included (1) "The personalized report was informative" and (2) "I learned something from the personalized report," and responses ranged from 1 (*strongly disagree*) to 5 (*strongly agree*). Items were averaged (Cronbach alpha=.84).

With respect to *perceived message quality*, participants were asked about their perceptions of the quality of the personalized report using a 5-item perceived message quality scale [54]. Items included statements such as "The personalized report was persuasive" and "I feel that the personalized report was convincing." Response options ranged from 1 (*strongly disagree*) to 5 (*strongly agree*) and were averaged across the 5 items (Cronbach alpha=.93).

Perceived trustworthiness was assessed with one item [38,39]: "I believed the information in the personalized report." Responses were on a 5-point scale from 1 (*strongly disagree*) to 5 (*strongly agree*).

To assess *perceived attractiveness*, participants were asked 1 item [40] on a 7-point scale (1=*very much* to 7=*not at all)*: "How attractive did you find the personalized report?"

For assessing *perceived message effectiveness*, a 3-item scale, adapted from Jensen et al [40] was used to ask participants about the persuasiveness of the personalized report. Questions asked (1) "Was the personalized report convincing?"; (2) Would people your age who smoke (who are not exercising regularly, who are not eating a healthy diet, and who are not at a healthy weight) be more likely to quit (to exercise regularly, to eat a healthier diet, and to aim for a healthy weight) after reading the personalized report?"; (3) "Would the personalized report be helpful in convincing your friends to quit smoking (to exercise regularly, to eat a healthy diet, and to aim for a healthy weight)?" Responses options ranged from 1 (*definitely no)* to 4 (*definitely yes*) and were averaged (Cronbach alpha=.89).

Engagement With Health Assessment Website

We adapted the 9-item Website Evaluation Questionnaire [55], originally developed to measure self-reported engagement, to ask participants about their evaluation of the overall CHART website. Responses were on a 5-point scale (1=*strongly disagree* to 5=*strongly agree*) and were averaged across three different items to derive three subscales. *Perceptions of personal relevance* assessed the degree to which participants felt the website was tailored (eg, "The information and advice provided by the website was appropriate for me"; Cronbach alpha=.85). The *perceptions of self-assessment and goal setting* subscale assessed the degree to which participants felt that the website helped them to reflect on their current behaviors and set goals (eg, "The website helped me to plan"; Cronbach alpha=.86), whereas the *engagement* subscale assessed the degree to which participants felt the website was attractive and enjoyable to use (eg, "The website was engaging"; Cronbach alpha=.89.).

For *satisfaction*, a single item asked participants the following: "How do you assess your participation in the online health assessment website in general?" Response options were on a 5-point scale and included 1 (*poor*), 2 (*average*), 3 (*good*), 4 (*very good)*, and 5 (*excellent*).

Demographic Characteristics

Participants reported age, sex, race, ethnicity, educational attainment, marital status, annual income, employment, and health insurance status. Data were collected through the CHART demographics module.

Statistical Analyses

Data were examined for outliers and distributions. Given that the distribution of data relating to the tobacco module was markedly different from the other behaviors (eating habits, PA, and weight management), we analyzed the data related to

tobacco separately from the other behaviors, whereas data on the three other behaviors were combined. In examining data, the distribution of the primary outcome variable (behavioral intentions) was similar across the PA, eating behavior, and weight modules, whereas it differed for the tobacco module. Consistent with a previous approach used to analyze data related to CHART [10], we collapsed data across the PA, eating behavior, and weight modules. Furthermore, as the tobacco module focused on an addictive behavior with a recommendation to quit and the other three behavioral modules were similar with respect to recommendations promoting adoption of behaviors, we anticipated that the messages might have similar effects across the three nontobacco behaviors.

We conducted multivariate analysis of variance to evaluate the main effect of each condition (ie, difference between mean response at one level of factor and mean response at other level, collapsing over the levels of all remaining factors) and interactions between conditions on our primary outcomes of interest (behavioral intentions and self-efficacy). Each model included the three experimental conditions, two-way and three-way interaction terms (expectancy priming x autonomy support, expectancy priming x exemplar, autonomy support x exemplar, and expectancy priming x autonomy support x exemplar), an intercept, and the grand mean-centered baseline measure of the outcome of interest as a covariate.

For models related to the health behaviors other than tobacco, the assigned health behavior module was also included as a covariate. Estimated marginal means based on models are reported. We used a similar approach to examine the effect of experimental conditions on measures of engagement and message perceptions. As these measures were collected only in the follow-up questionnaire, analyses did not control for a baseline measure. All analyses were conducted using Statistical Package for the Social Sciences (SPSS) version 23.0 (IBM Corp).

Results

Participants

Of 3163 respondents in MTurk, 2838 completed the online screener and were randomized to 1 of 32 possible conditions (8 conditions across 4 behaviors). Among 2749 individuals who completed the baseline questionnaire, 2285 proceeded to take the CHART assessment. Upon reviewing their personalized reports, 1971 participants completed the follow-up questionnaire. Due to an error with skip patterns that resulted in missing responses related to self-efficacy, 155 participants in the tobacco module were excluded from analyses. An additional 26 participants were excluded because their website activity indicated cases with duplicate IDs and unpaired assessments because of technical issues, which resulted in 1945 participants used in analyses. Figure 1 and Table 1 show the number of participants analyzed by health behavior module and condition, respectively. Characteristics of the 1945 participants are summarized in Table 2. Participants (N=1945) were on average aged 36 years, with the majority being female (54.6%, 1062/1945), married (54.6%, 1062/1945), and college graduates (52.4%, 1019/1945).

Effects on Primary Outcomes

Self-Efficacy

For tobacco-related self-efficacy, analyses revealed a significant main effect of the exemplar condition, $F_{1,266}$=4.157, P=.04, η^2=0.015. Participants who received exemplar messages (mean 3.31, SE 0.060) rated their self-efficacy to quit tobacco higher than those who did not receive exemplar messages (mean 3.14, SE 0.057). Expectancy priming ($F_{1,266}$=0.836, P=.36) and autonomy support ($F_{1,266}$=0.019, P=.89) conditions did not have main effects on tobacco-related self-efficacy.

The three-way interaction between the conditions was statistically significant ($F_{1,266}$=5.807, P=.02, η^2=0.021) and is illustrated in Figure 2. The strength of the effect of the exemplar condition was moderated by the other conditions. For those in the tailored priming condition, mean self-efficacy was highest among those who received the exemplar message with no autonomy support (mean 3.47, SE 0.129). The next highest self-efficacy scores were among those in the standard priming condition, who received both the exemplar and autonomy supportive messages (mean 3.35, SE 0.111).

Results on the health behaviors other than tobacco showed no significant main effects of the three experimental conditions on self-efficacy at follow-up (priming: $F_{1,1501}$=0.518, P=.47; autonomy support: $F_{1,1501}$=0.165, P=.685; and exemplar: $F_{1,1501}$=0.695, P=.41).

Table 1. Experimental conditions and cell sizes by health behavior module.

Group	Experimental condition			Health behavior module (N)			
	Tailored expectancy priming	Autonomy support	Exemplar	Tobacco	Physical activity	Eating habits	Weight
1	No	No	No	56	71	62	69
2	No	No	Yes	57	62	62	61
3	No	Yes	No	42	64	67	63
4	No	Yes	Yes	57	66	63	60
5	Yes	No	No	58	59	78	61
6	Yes	No	Yes	52	59	62	57
7	Yes	Yes	No	54	59	61	63
8	Yes	Yes	Yes	55	61	66	58

Table 2. Characteristics of participants (N=1945) in experiments assessing Carolina Health Assessment and Resource Tool (CHART) personalized reports.

Characteristic	Value
Age in years, mean (SD)	36.22 (11.01)
Sex, n (%)	
Female	1062 (54.60)
Male	883 (45.40)
Race[a] , n (%)	
Non-Hispanic white	1675 (86.12)
African American	148 (7.61)
Asian	134 (6.89)
American Indian or native American	41 (2.11)
Pacific Islander	4 (0.21)
Other	21 (1.08)
Ethnicity, n (%)	
Hispanic	128 (6.58)
Marital status, n (%)	
Married or living as married	1062 (54.60)
Not married	883 (45.40)
Education, n (%)	
≤High school	230 (11.83)
Some college or technical school	696 (35.78)
≥College graduate	1019 (52.39)
Annual income (USD), n (%)	
<$35,000	792 (40.72)
$35,000 to <$75,000	727 (37.38)
≥$75,000	379 (19.49)
Employment status, n (%)	
Employed	1385 (71.21)
Not employed	552 (28.38)
Have health insurance, n (%)	1564 (80.41)

[a]Participants could choose all that apply.

The two-way interaction effect of priming and autonomy support conditions was significant ($F_{1,1501}$=7.028, P=.008, η^2=0.005), such that the effect of autonomy support was higher within the standard priming condition (mean 3.28, SE 0.036) than the tailored priming condition (mean 3.16, SE 0.037). As with self-efficacy related to tobacco, the three-way interaction among the conditions had a significant effect on self-efficacy related to the other health behaviors ($F_{1,1501}$=3.925, P=.048, η^2=0.003).

Figure 2 shows the effect of autonomy support and exemplar conditions by priming condition. The exemplar condition increased the effect of autonomy support within the standard priming condition, such that the highest self-efficacy scores were reported among those who were in the standard priming condition and received both autonomy and exemplar messages (mean 3.33, SE 0.052). Within the tailored priming condition, mean self-efficacy scores were lowest among those in the autonomy support condition, with (mean 3.14, SE 0.052) or without an exemplar message (mean 3.18, SE 0.053).

In analyses of self-efficacy by individual health behaviors, we found that the significant interaction effect of expectancy priming and autonomous support was specific to those within the PA module ($F_{1,491}$=7.185, P=.008, η^2=0.014). Figure 3 illustrates the two-way interaction, such that autonomy support had stronger effects on self-efficacy for PA in the standard priming condition (mean 3.47, SE 0.070), whereas messages without autonomy support had stronger effects in the tailored priming condition (mean 3.36, SE 0.074).

Figure 2. Estimated means (SE) for self-efficacy at follow-up as a function of three-way interaction of expectancy priming, autonomy support, and exemplar conditions. Error bars are SEs of the means. Higher scores represent higher self-efficacy. Tobacco (top): three-way interaction effect (P=.02) of autonomy support and exemplar conditions on self-efficacy to quit smoking, by priming condition. Physical activity, eating habits, weight (bottom): three-way interaction effect (P=.048) of autonomy support and exemplar conditions on self-efficacy to engage in physical activity, eating habits, and weight management behaviors, by priming condition.

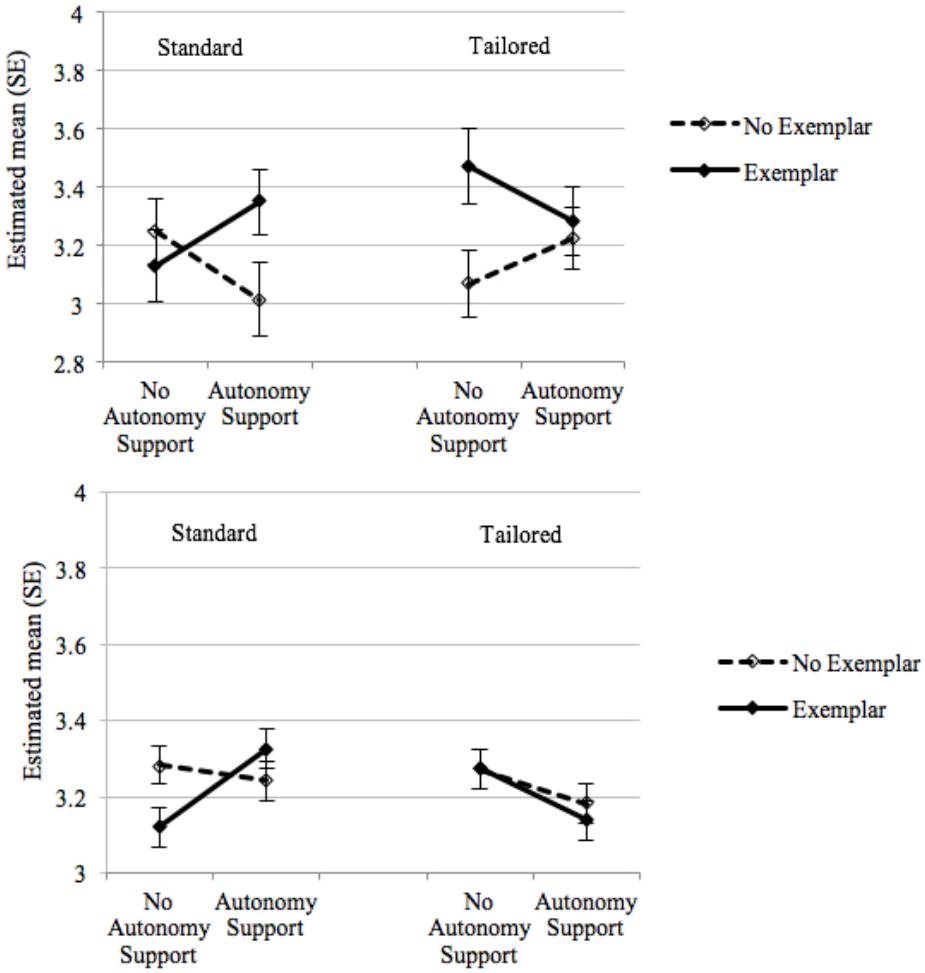


Figure 3. Estimated means (SE) for self-efficacy for physical activity at follow-up as a function of two-way interaction of expectancy priming and autonomy support. Error bars are SEs of the means. Higher scores represent higher self-efficacy.

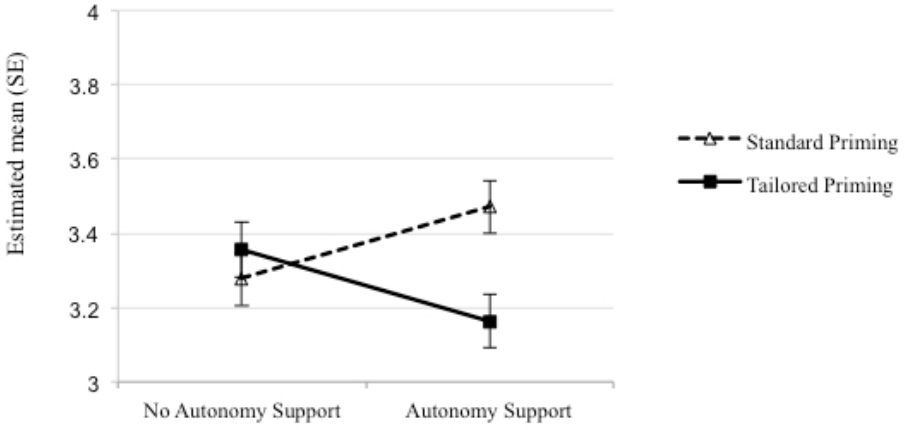
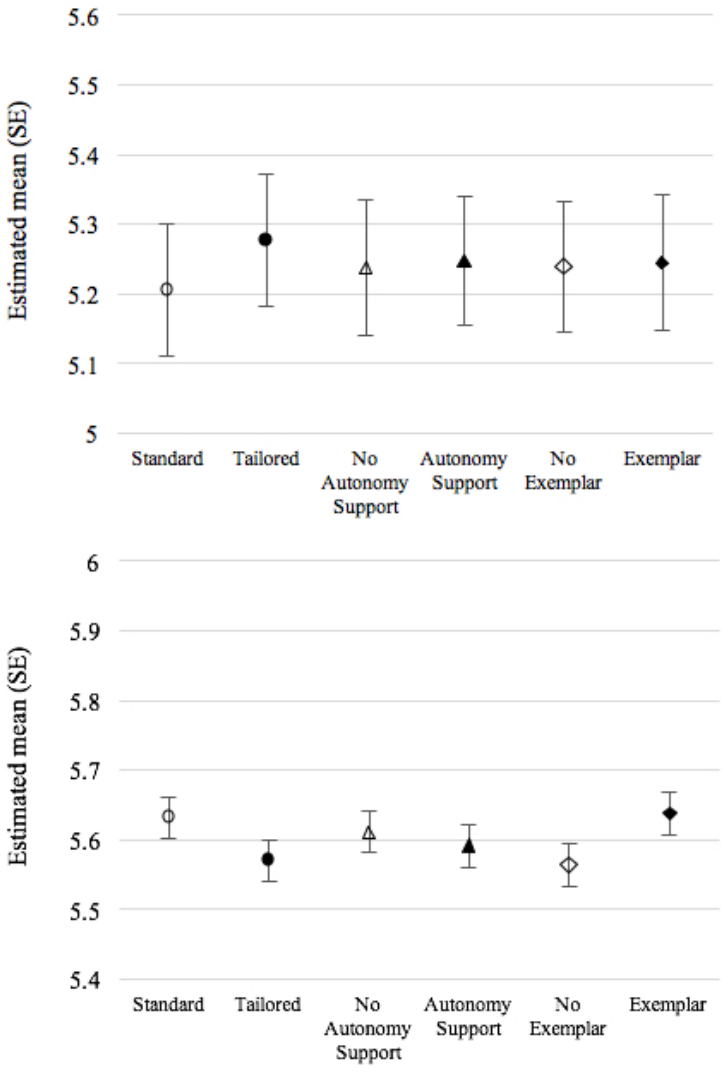


Figure 4. Estimated means (SE) for behavioral intentions at follow-up as a function of main effects of experimental conditions. Error bars are SEs of the means. Higher scores represent higher behavioral intentions. Tobacco (top): priming (*P*=.59), autonomy support (*P*=.94), and exemplar (*P*=.97) effects on behavioral intentions to quit smoking, controlling for baseline intention scores. Physical activity, eating habits, weight (bottom): priming (*P*=.15), autonomy support (*P*=.64), and exemplar (*P*=.08) effects on behavioral intentions to engage in other health behaviors, controlling for baseline intention scores.



Behavioral Intentions

Figure 4 shows the results for behavioral intentions to quit smoking. There were no significant main effects of the three conditions. Behavioral intentions among participants that received the standard priming message (mean 5.21, SE 0.095) did not differ from those who received the tailored priming message (mean 5.28, SE 0.095; $F_{1,267}$=0.292, P=.59). Similarly, there was no main effect for autonomy support (mean 5.25, SE 0.093) versus no autonomy support (mean 5.24, SE 0.097; $F_{1,267}$=0.005, P=.94), or the exemplar (mean 5.24, SE 0.097) versus no exemplar (mean 5.24, SE 0.094; $F_{1,267}$=0.001, P=.97). No significant interactions were found among the experimental conditions.

For the other health behaviors (Figure 4), the main effect of exemplar messages on behavioral intentions was in the hypothesized direction but did not reach statistical significance ($F_{1,1503}$=3.026, P=.08). Mean behavioral intention scores were 5.64 (SE 0.031) for those who received exemplar messages and 5.56 (SE 0.030) for those who did not. There were no main effects of priming or autonomy support. Mean behavioral intention scores were similar between standard priming (mean 5.63, SE 0.030) and tailored priming conditions (mean 5.57, SE 0.030; $F_{1,1503}$=2.06, P=.15) and between the autonomy support (mean 5.59, SE 0.030) and no autonomy support (mean 5.61, SE 0.030; $F_{1,1503}$=0.223, P=.64) conditions. There were no significant interactions between any of the experimental conditions.

Table 3. Engagement and perceptions of the Carolina Health Assessment and Resource Tool (CHART) personalized report and website at follow-up (tobacco module).

Scale	Scale range	Tobacco, mean (SD)					
		Standard priming (N=200)	Tailored priming (N=211)	No autonomy support (N=213)	Autonomy support (N=198)	No exemplar (N=202)	Exemplar (N=209)
Perceived message relevance	1-5	3.58 (0.91)	3.64 (0.90)	3.59 (0.92)	3.62 (0.90)	3.59 (0.89)	3.62 (0.93)
Perceived informativeness	1-5	3.69 (0.88)	3.74 (0.97)	3.72 (0.91)	3.71 (0.94)	3.71 (0.95)	3.73 (0.91)
Perceived quality	1-5	3.63 (0.87)	3.63 (0.92)	3.64 (0.90)	3.62 (0.88)	3.63 (0.88)	3.63 (0.90)
Perceived trustworthiness	1-5	3.94 (0.85)	4.01 (0.89)	3.97 (0.91)	3.99 (0.82)	3.93 (0.89)	4.02 (0.85)
Perceived attractiveness	1-7	3.97 (1.66)	3.89 (1.76)	3.80 (1.74)	4.06 (1.68)	3.95 (1.71)	3.91 (1.72)
Perceived message effectiveness	1-4	2.73 (0.63)	2.77 (0.65)	2.74 (0.62)	2.75 (0.65)	2.72 (0.66)	2.78 (0.62)
Perceptions of personal relevance	1-5	3.69 (0.76)	3.70 (0.83)	3.66 (0.82)	3.73 (0.78)	3.68 (0.80)	3.71 (0.79)
Perceptions of self-assessment and goal setting	1-5	3.47 (0.91)	3.53 (0.93)	3.47 (0.90)	3.54 (0.93)	3.45 (0.91)	3.56 (0.92)
Engagement	1-5	3.57 (0.89)	3.62 (0.96)	3.58 (0.90)	3.62 (0.94)	3.48 (0.96)	3.72 (0.87)
Participation in CHART[a]	1-5	3.68 (0.91)	3.67 (0.93)	3.68 (0.89)	3.67 (0.95)	3.57 (0.96)	3.78 (0.87)

[a]CHART: Carolina Health Assessment and Resource Tool.

Table 4. Engagement and perceptions of the Carolina Health Assessment and Resource Tool (CHART) personalized report and website at follow-up (physical activity, eating behaviors, and weight modules).

Scale	Scale range	Physical activity, eating behaviors, and weight, mean (SD)					
		Standard priming (N=744)	Tailored priming (N=715)	No autonomy support (N=743)	Autonomy support (N=716)	No exemplar (N=744)	Exemplar (N=715)
Perceived message relevance	1-5	3.57 (0.88)	3.56 (0.92)	3.59 (0.89)	3.54 (0.89)	3.59 (0.89)	3.55 (0.92)
Perceived informativeness	1-5	3.75 (0.94)	3.72 (0.94)	3.70 (0.96)	3.76 (0.92)	3.73 (0.94)	3.74 (0.94)
Perceived quality	1-5	3.63 (0.88)	3.62 (0.90)	3.61 (0.89)	3.64 (0.88)	3.63 (0.88)	3.61 (0.89)
Perceived trustworthiness	1-5	3.96 (0.85)	3.95 (0.86)	3.96 (0.88)	3.95 (0.83)	3.93 (0.88)	3.97 (0.83)
Perceived attractiveness	1-7	3.89 (1.69)	3.89 (1.68)	3.93 (1.70)	3.85 (1.67)	3.90 (1.72)	3.88 (1.72)
Perceived message effectiveness	1-4	2.83 (0.61)	2.81 (0.62)	2.82 (0.62)	2.82 (0.61)	2.81 (0.62)	2.83 (0.62)
Perceptions of personal relevance	1-5	3.64 (0.81)	3.65 (0.84)	3.64 (0.83)	3.66 (0.82)	3.67 (0.80)	3.62 (0.85)
Perceptions of self-assessment and goal setting	1-5	3.54 (0.89)	3.55 (0.89)	3.52 (0.89)	3.57 (0.90)	3.57 (0.86)	3.52 (0.93)
Engagement	1-5	3.65 (0.89)	3.63 (0.88)	3.64 (0.88)	3.64 (0.90)	3.65 (0.89)	3.63 (0.89)
Participation in CHART[a]	1-5	3.63 (0.96)	3.67 (0.93)	3.65 (0.93)	3.65 (0.97)	3.66 (0.95)	3.63 (0.95)

[a]CHART: Carolina Health Assessment and Resource Tool.

Effects on User Engagement and Message Perceptions

Tables 3 and 4 show the mean scores related to engagement and perceptions of the personalized report and CHART website by main effects of the experimental conditions. Overall, participants reported positive perceptions of the tailored feedback reports and CHART website. Mean ratings of the personalized report regarding perceived message relevance, informativeness, message quality, and trustworthiness (1=*strongly disagree* to 5=*strongly agree*) ranged from 3.58 (SD 0.91) to 4.02 (SD 0.85) for the tobacco module (Table 3) and from 3.54 (SD 0.89) to 3.97 (SD 0.83) for the nontobacco modules (Table 4). Participants across both the tobacco-related and nontobacco modules reported average scores for attractiveness of the personalized reports (1=*very much* to 7=*not at all*), ranging from 3.80 (SD 1.74) to 4.06 (SD 1.68) and 3.85 (SD 1.67) to 3.93 (SD 1.70), respectively.

Message effectiveness of personalized reports (1=*definitely no* to 4=*definitely yes*) was rated more positively for both the tobacco module and other behaviors. Evaluations of the overall CHART website (1=*strongly disagree* to 5=*strongly agree*) were generally positive across all three subscales (personal relevance, perceptions of self-assessment and goal setting, and engagement). Scores ranged from 3.45 (SD 0.91) to 3.78 (SD 0.87) for the tobacco modules (Table 3) and 3.52 (SD 0.93) to 3.67 (SD 0.80) for the nontobacco modules (Table 4). Satisfaction ratings regarding participation in CHART fell between *good* to *very good* across all modules. There were no significant differences in engagement and perceptions among groups.

Discussion

Principal Findings

Findings from this randomized factorial experiment showed that tailored priming before presentation of a tailored feedback report, with use of an exemplar to model smoking cessation behavior, produced the largest effect on self-efficacy in the tobacco module. For the other three modules (PA, eating habits, and weight), self-efficacy was highest among those who received standard priming of the feedback and messages offering both autonomy support and an exemplar. Messages featuring tailored expectancy priming, autonomy support, or exemplars did not improve behavioral intentions to engage in healthy behaviors or result in differences in engagement and message perceptions. These empirical findings on various theory-driven messages delivered in response to an online health risk assessment contribute to the relatively sparse literature guiding the optimal presentation of online tailored feedback to individuals. Overall, results of this study suggest that using exemplars in tailored feedback messages has the potential to improve self-efficacy in the early phases of behavior change interventions. Furthermore, the addition of standard priming messages before presentation of feedback, along with autonomy supportive messages, could help optimize message effects on self-efficacy.

Comparison With Prior Work

Self-Efficacy

Exemplars

In this study, using an exemplar showed potential for improving self-efficacy across all of the behaviors (tobacco, PA, eating habits, and weight management). Among participants in the tobacco module, messages with exemplars produced the highest self-efficacy scores. Strecher et al [36] previously demonstrated that high-depth tailored (ie, tailored to several characteristics

beyond name and gender) success stories delivered through a Web-based smoking cessation program were effective for improving smoking abstinence at 6 months. Similarly, Sarge and Knobloch-Westerwick [56] showed that using exemplars in an online health article that modeled successful weight loss behavior improved weight loss self-efficacy. The positive effects on self-efficacy observed among those receiving exemplar messages are consistent with SCT and strategies such as observational learning, vicarious experience, and verbal persuasion that are theorized to promote self-efficacy [32,34,35].

Expectancy Priming

Interestingly, the highest self-efficacy scores among participants in the tobacco module were observed in those who received exemplar messages along with tailored priming, whereas the next highest were among those receiving exemplar messages with standard priming and autonomy supportive messages. Likewise, for the other health behavior modules (PA, eating habits, and weight), messages that included standard priming with autonomy supportive and exemplar messages resulted in the highest self-efficacy scores. The potentially positive effects of priming are consistent with work by Webb et al [19], which showed that pretreatment expectancy priming (both standard and tailored) of smoking cessation materials resulted in improvements in readiness to quit smoking and smoking-related knowledge, although not self-efficacy. Another study showed that patients who were primed with physician advice before receiving printed health education materials were more likely than those who did not receive physician advice, to report changes in diet and PA and attempt to quit smoking [57]. Our findings suggest that the inclusion of expectancy priming, such as making it clear that a message is personally tailored for that individual, before delivery of personalized feedback may improve the effects of tailored messages in the context of online health risk assessment tools.

Autonomy Support

The interaction between the effects of standard priming and autonomy supportive messages on self-efficacy (ie, the effect of autonomy support was higher in the standard priming condition) demonstrates potential for further study. It is unclear why self-efficacy would be higher among those receiving the standard priming and autonomy supportive messages. It is possible that the standard priming influenced individuals' positive expectancies related to standard health messages, and viewing subsequent autonomy supportive messages matched expectations that the feedback report would be more general in nature, as the messages offered various options and encouraged individuals to reflect on their own preferences. Previous research has shown that among participants receiving three tailored newsletters aimed at improving autonomous motivation, those who preferred and received more autonomy supportive communication increased their fruit and vegetable intake relative to those who received tailored newsletters not focused on autonomous motivation [12]. Few research studies have examined the effectiveness of tailoring online health communication messages based on individuals' need or preference for autonomy [58]. Future work examining the mechanisms of interaction between expectancy priming and messages using either autonomy supportive or exemplar messages appear warranted. In particular, identifying and assessing preferences for autonomy supportive communication, or more narrative forms that include behavioral models, may help improve the development of more personally relevant messaging in tailored feedback communications. Given the scarcity of research that has evaluated online health communications that tailor message framing to match an individuals' need for autonomy and other information processing styles [58], there is a need to further elucidate the effectiveness of tailoring messages based on autonomy supportive preferences. Studies that examine the effects of tailoring on need for autonomy alone or in combination with other psychosocial constructs could advance our understanding of the potential to improve the effectiveness of online health communications that are tailored to individuals' information processing preferences.

Behavioral Intentions

Although we expected to observe improvements in behavioral intentions as a result of the various message enhancements, this was not borne out in our findings. In the nontobacco modules (PA, eating habits, and weight), there was a nonsignificant trend for messages with exemplars to improve behavioral intentions to engage in the behaviors. This lack of effect on behavioral intentions is similar to findings from other message testing studies that have sought to improve intentions for various health behaviors using a one-time delivery of targeted or framed messages [59-61]. In the context of a computer-tailored nutrition intervention, Oenema et al [62] showed that the tailored intervention improved intentions to change vegetable consumption relative to generic nutrition information and no-information control groups. This effect was mediated by perceived message relevance and perceived individualization. Given that the existing CHART messages were already tailored to individuals' current behaviors and other psychosocial factors and aimed at improving behavioral intentions, it is possible there was a ceiling effect with little room for improvement beyond that produced by existing messages. Furthermore, the nonsignificant differences in behavioral intentions may be attributed to the lack of differences in perceived message relevance among the three message factors. Although behavioral intentions is a common construct in health behavior theories, studies highlight the gap between behavioral intention and subsequent behavior [63] and have shown that people's expectations about what they will do are more predictive of subsequent behavior than their intentions [64]. Future studies of the effectiveness of tailored feedback messages might consider alternative outcome measures that have demonstrated better predictive validity of behavior, such as expectations, and examine the effects of multiple or frequent feedback reports delivered over time.

User Engagement and Message Perceptions

The comparability across the message factors with respect to perceptions and engagement with the personalized reports and CHART website suggests that the message conditions were equally appealing, relevant, and engaging. Previous studies have found that a variety of message perceptions (eg, relevance, persuasiveness, importance, and helpfulness) have mediated the relationship between tailored messaging and behaviors or

behavioral intentions [38-40,62,65]. For example, perceived message relevance has been shown to be a mediator of the positive effects of tailored messaging on fruit and vegetable intake [38,39], vegetable intake intentions [62], and breast screening intentions [40], indicating the existence of significant relationships between the tailored materials and perceived message relevance.

In this study, we did not find such associations between the tailoring enhancements and message perceptions. Given that all experimental conditions provided tailored feedback messages, it is not surprising that there was little variation in perceptions and user engagement among the message factors. The relatively slight differences in phrasing of messages may not have been sufficient to influence various message perceptions as overall content and suggestions for making healthy behavior changes were consistent across conditions. The ratings for engagement with and perceptions of the CHART website indicate that participants found the overall health assessment website with feedback to be engaging and personally relevant. Participant ratings in this study were comparable with findings on self-reported engagement in a study that compared the effects of Internet-delivered assessment with and without tailored feedback versus generic information on self-management of bowel problems [55]. A recent systematic review provided an integrative definition of engagement with online behavior change interventions, which incorporates both subjective experiences and extent of usage, and offered a conceptual framework to guide measurement and evaluation of the relationship between engagement and intervention effectiveness [16]. Future research on tailored personalized feedback could incorporate both subjective and objective measurements of engagement to identify specific dimensions of engagement that have a greater influence on the effectiveness of health messages. Further elucidation of whether engagement may mediate or moderate the relationship between tailored feedback messages and intended behavioral outcomes is necessary.

Limitations

While this study used an innovative strategy and factorial design to efficiently identify the most impactful message conditions to optimize tailored feedback reports, several limitations should be considered. Participants completed the pre- and postfeedback assessments during a single occasion, so the persistence of findings over time is unknown.

Although improving self-efficacy for behavior change and behavioral intentions may require more than a one-time administration of a health risk assessment with tailored feedback, our study was an experiment designed to test the effects of different message features on psychosocial factors and not an intervention for behavior change. Our findings may have implications for creating more impactful messages within the context of behavior change interventions. The sample recruited through MTurk was relatively homogeneous with respect to age, race, ethnicity, and education, which limits generalizability of our findings to a more diverse population. However, this recruitment approach facilitated rapid and efficient testing of multiple message conditions over a short period of time, and results are useful for generating hypotheses to be tested in future alternate populations that are harder and most costly to recruit. As all measures and outcomes were self-reported, over- or underreporting and responses influenced by social desirability may have biased our results, though presumably randomization would have evenly distributed any biased reporting across conditions.

Another limitation was the lack of an experimental condition without priming. Although this precluded us from evaluating the effects of any priming (whether standard or tailored) compared with none, we observed an interesting interaction effect, whereby combining standard priming with autonomy supportive messages resulted in higher self-efficacy scores. Finally, we observed relatively small effects of the message manipulations on our outcomes of interest. To minimize participant burden, we opted to use a single-item to assess self-efficacy, a limitation that may have led to insufficient variance to examine group differences. Although participants' mean intentions and self-efficacy scores were relatively high at baseline, which possibly resulted in a ceiling effect and diminished our ability to observe associations, the large sample size was adequately powered to detect some differential effects among the message conditions.

Conclusions

Overall, our study findings among MTurk workers suggest message characteristics that have the potential to enhance message impact on self-efficacy. In the context of an online health behavior assessment tool, the use of exemplars to convey tailored feedback may help promote improvements in self-efficacy related to tobacco cessation, PA, eating habits, and weight control. As findings among MTurk workers may not generalize to others who are seeking behavioral interventions, further evaluation of whether exemplars, priming, and autonomy supportive messages can enhance the impact of tailored feedback on cancer prevention–related behaviors among other populations is warranted.

Acknowledgments

This research was supported by the University Cancer Research Fund. For graphic design and software development, the authors thank UNC CHAI Core, which is supported in part by a grant from the National Institutes of Health (DK056350) to the University of North Carolina Nutrition Obesity Research Center and from NCI (P30-CA16086) to the Lineberger Comprehensive Cancer Center. The authors are grateful to Anne Cabell, MPH, Jessica Zahurak, and Hannah Lerner, MSPH for their assistance with the study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Examples of experimental stimuli in personalized reports by condition.

[PDF File (Adobe PDF File), 373KB - jmir_v20i3e63_app1.pdf]

References

1. Soler RE, Leeks KD, Razi S, Hopkins DP, Griffith M, Aten A, et al. Task Force on Community Preventive Services. A systematic review of selected interventions for worksite health promotion. The assessment of health risks with feedback. Am J Prev Med 2010 Feb;38(2 Suppl):S237-S262. [doi: 10.1016/j.amepre.2009.10.030] [Medline: 20117610]
2. Noar SM, Benac CN, Harris MS. Does tailoring matter? Meta-analytic review of tailored print health behavior change interventions. Psychol Bull 2007 Jul;133(4):673-693. [doi: 10.1037/0033-2909.133.4.673] [Medline: 17592961]
3. 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]
4. 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]
5. Brug J, Campbell M, van Assema P. The application and impact of computer-generated personalized nutrition education: a review of the literature. Patient Educ Couns 1999 Feb;36(2):145-156. [Medline: 10223019]
6. Kreuter MW, Farrell DW, Olevitch LR, Brennan LK. Tailoring Health Messages: Customizing Communication With Computer Technology. New York, NY: Routledge; 2013.
7. Velicer WF, Prochaska JO, Redding CA. Tailored communications for smoking cessation: past successes and future directions. Drug Alcohol Rev 2006 Jan;25(1):49-57. [doi: 10.1080/09595230500459511] [Medline: 16492577]
8. 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;3(2):e42 [FREE Full text] [doi: 10.2196/mhealth.4160] [Medline: 25977197]
9. Harrington NG, Noar SM. Reporting standards for studies of tailored interventions. Health Educ Res 2012 Apr;27(2):331-342 [FREE Full text] [doi: 10.1093/her/cyr108] [Medline: 22156230]
10. Comello ML, Qian X, Deal AM, Ribisl KM, Linnan LA, Tate DF. Impact of game-inspired infographics on user engagement and information processing in an eHealth program. J Med Internet Res 2016 Sep 22;18(9):e237 [FREE Full text] [doi: 10.2196/jmir.5976] [Medline: 27658469]
11. Brug J, Oenema A, Campbell M. Past, present, and future of computer-tailored nutrition education. Am J Clin Nutr 2003 Apr;77(4 Suppl):1028S-1034S [FREE Full text] [Medline: 12663313]
12. Resnicow K, Davis RE, Zhang G, Konkel J, Strecher VJ, Shaikh AR, et al. Tailoring a fruit and vegetable intervention on novel motivational constructs: results of a randomized study. Ann Behav Med 2008 Apr;35(2):159-169. [doi: 10.1007/s12160-008-9028-9] [Medline: 18401673]
13. Collins LM, Trail JB, Kugler KC, Baker TB, Piper ME, Mermelstein RJ. Evaluating individual intervention components: making decisions based on the results of a factorial screening experiment. Transl Behav Med 2014 Sep;4(3):238-251 [FREE Full text] [doi: 10.1007/s13142-013-0239-7] [Medline: 25264464]
14. Collins LM, Murphy SA, Nair VN, Strecher VJ. A strategy for optimizing and evaluating behavioral interventions. Ann Behav Med 2005 Aug;30(1):65-73. [doi: 10.1207/s15324796abm3001_8] [Medline: 16097907]
15. Collins LM, Murphy SA, Strecher V. The multiphase optimization strategy (MOST) and the sequential multiple assignment randomized trial (SMART): new methods for more potent eHealth interventions. Am J Prev Med 2007 May;32(5 Suppl):S112-S118 [FREE Full text] [doi: 10.1016/j.amepre.2007.01.022] [Medline: 17466815]
16. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
17. Schwarzer R, Luszczynska A. 2008. Perceived self-efficacy URL: https://cancercontrol.cancer.gov/brp/research/constructs/self_efficacy.html [accessed 2017-12-10] [WebCite Cache ID 6vUSVM7DH]
18. Hagger MS, Chatzisarantis NLD, Biddle SJH. The influence of autonomous and controlling motives on physical activity intentions within the theory of planned behaviour. Br J Health Psychol 2002 Sep;7(Part 3):283-297. [doi: 10.1348/135910702760213689] [Medline: 12614501]
19. Webb MS, Hendricks PS, Brandon TH. Expectancy priming of smoking cessation messages enhances the placebo effect of tailored interventions. Health Psychol 2007 Sep;26(5):598-609. [doi: 10.1037/0278-6133.26.5.598] [Medline: 17845111]

20. Webb MS, Simmons VN, Brandon TH. Tailored interventions for motivating smoking cessation: using placebo tailoring to examine the influence of expectancies and personalization. Health Psychol 2005 Mar;24(2):179-188. [doi: 10.1037/0278-6133.24.2.179] [Medline: 15755232]
21. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [Medline: 11392867]
22. Teixeira PJ, Carraça EV, Markland D, Silva MN, Ryan RM. Exercise, physical activity, and self-determination theory: a systematic review. Int J Behav Nutr Phys Act 2012;9:78 [FREE Full text] [doi: 10.1186/1479-5868-9-78] [Medline: 22726453]
23. Resnicow K, Jackson A, Blissett D, Wang T, McCarty F, Rahotep S, et al. Results of the healthy body healthy spirit trial. Health Psychol 2005 Jul;24(4):339-348. [doi: 10.1037/0278-6133.24.4.339] [Medline: 16045368]
24. Crane MM, Lutes LD, Ward DS, Bowling JM, Tate DF. A randomized trial testing the efficacy of a novel approach to weight loss among men with overweight and obesity. Obesity (Silver Spring) 2015 Dec;23(12):2398-2405 [FREE Full text] [doi: 10.1002/oby.21265] [Medline: 26727117]
25. Deci EL, Ryan RM. The "what" and "why" of goal pursuits: human needs and the self-determination of behavior. Psychol Inq 2000 Oct;11(4):227-268. [doi: 10.1207/s15327965pli1104_01]
26. Deci EL, Ryan RM. Facilitating optimal motivation and psychological well-being across life's domains. Canadian Psychol 2008;49(1):14-23. [doi: 10.1037/0708-5591.49.1.14]
27. Ng JY, Ntoumanis N, Thøgersen-Ntoumani C, Deci EL, Ryan RM, Duda JL, et al. Self-determination theory applied to health contexts: a meta-analysis. Perspect Psychol Sci 2012 Jul;7(4):325-340. [doi: 10.1177/1745691612447309] [Medline: 26168470]
28. Shaikh AR, Vinokur AD, Yaroch AL, Williams GC, Resnicow K. Direct and mediated effects of two theoretically based interventions to increase consumption of fruits and vegetables in the Healthy Body Healthy Spirit trial. Health Educ Behav 2011 Oct;38(5):492-501. [doi: 10.1177/1090198110384468] [Medline: 21596903]
29. Ryan RM, Patrick H, Deci EL, Williams GC. Facilitating health behaviour change and its maintenance: interventions based on self-determination theory. Eur Health Psychol 2008;10(1):2-5 [FREE Full text]
30. Patrick H, Williams GC. Self-determination theory: its application to health behavior and complementarity with motivational interviewing. Int J Behav Nutr Phys Act 2012;9:18 [FREE Full text] [doi: 10.1186/1479-5868-9-18] [Medline: 22385676]
31. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977 Mar;84(2):191-215. [Medline: 847061]
32. Bandura A. Self-efficacy mechanism in human agency. Am Psychol 1982;37(2):122-147. [doi: 10.1037/0003-066X.37.2.122]
33. Zillmann D. Exemplification theory: Judging the whole by some of its parts. Media Psychol 1999 Mar;1(1):69-94. [doi: 10.1207/s1532785xmep0101_5]
34. Bandura A. Social cognitive theory: an agentic perspective. Annu Rev Psychol 2001;52:1-26. [doi: 10.1146/annurev.psych.52.1.1] [Medline: 11148297]
35. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
36. Strecher VJ, McClure JB, Alexander GL, Chakraborty B, Nair VN, Konkel JM, et al. Web-based smoking-cessation programs: results of a randomized trial. Am J Prev Med 2008 May;34(5):373-381 [FREE Full text] [doi: 10.1016/j.amepre.2007.12.024] [Medline: 18407003]
37. Knobloch-Westerwick S, Sarge MA. Impacts of exemplification and efficacy as characteristics of an online weight-loss message on selective exposure and subsequent weight-loss behavior. Commun Res 2015 Jun;42(4):547-568. [doi: 10.1177/0093650213478440]
38. Ko LK, Campbell MK, Lewis MA, Earp JA, Devellis B. Information processes mediate the effect of a health communication intervention on fruit and vegetable consumption. J Health Commun 2011 Mar;16(3):282-299 [FREE Full text] [doi: 10.1080/10810730.2010.532294] [Medline: 21132593]
39. Ko LK, Campbell MK, Lewis MA, Earp J, Devellis B. Mediators of fruit and vegetable consumption among colorectal cancer survivors. J Cancer Surviv 2010 Jun;4(2):149-158 [FREE Full text] [doi: 10.1007/s11764-010-0116-z] [Medline: 20186574]
40. Jensen JD, King AJ, Carcioppolo N, Davis L. Why are tailored messages more effective? A multiple mediation analysis of a breast cancer screening intervention. J Commun 2012 Oct;62(5):851-868 [FREE Full text] [doi: 10.1111/j.1460-2466.2012.01668.x] [Medline: 26405350]
41. Mason W, Suri S. Conducting behavioral research on Amazon's Mechanical Turk. Behav Res Methods 2012 Mar;44(1):1-23. [doi: 10.3758/s13428-011-0124-6] [Medline: 21717266]
42. Carter RR, DiFeo A, Bogie K, Zhang G, Sun J. Crowdsourcing awareness: exploration of the ovarian cancer knowledge gap through Amazon Mechanical Turk. PLoS One 2014;9(1):e85508 [FREE Full text] [doi: 10.1371/journal.pone.0085508] [Medline: 24465580]
43. Kaczynski AT, Wilhelm Stanis SA, Hipp JA. Point-of-decision prompts for increasing park-based physical activity: a crowdsource analysis. Prev Med 2014 Dec;69:87-89 [FREE Full text] [doi: 10.1016/j.ypmed.2014.08.029] [Medline: 25204987]

44. Gardner RM, Brown DL, Boice R. Using Amazon's Mechanical Turk website to measure accuracy of body size estimation and body dissatisfaction. Body Image 2012 Sep;9(4):532-534. [doi: 10.1016/j.bodyim.2012.06.006] [Medline: 22832086]
45. Hall MG, Ribisl KM, Brewer NT. Smokers' and nonsmokers' beliefs about harmful tobacco constituents: implications for FDA communication efforts. Nicotine Tob Res 2014 Mar;16(3):343-350 [FREE Full text] [doi: 10.1093/ntr/ntt158] [Medline: 24151139]
46. Baig SA, Byron MJ, Boynton MH, Brewer NT, Ribisl KM. Communicating about cigarette smoke constituents: an experimental comparison of two messaging strategies. J Behav Med 2016 Sep 23;40(2):352-359. [doi: 10.1007/s10865-016-9795-x] [Medline: 27663553]
47. UNC Lineberger Comprehensive Cancer Center. CHART URL: https://chart.unc.edu/index.html [accessed 2018-02-10] [WebCite Cache ID 6x7dHw5HB]
48. Emmons KM, Linnan LA, Shadel WG, Marcus B, Abrams DB. The Working Healthy Project: a worksite health-promotion trial targeting physical activity, diet, and smoking. J Occup Environ Med 1999 Jul;41(7):545-555. [Medline: 10412096]
49. Marcus BH, Selby VC, Niaura RS, Rossi JS. Self-efficacy and the stages of exercise behavior change. Res Q Exerc Sport 1992 Mar;63(1):60-66. [doi: 10.1080/02701367.1992.10607557] [Medline: 1574662]
50. Armitage CJ, Conner M. Distinguishing perceptions of control from self-efficacy: predicting consumption of a low-fat diet using the theory of planned behavior. J Appl Social Pyschol 1999 Jan;29(1):72-90. [doi: 10.1111/j.1559-1816.1999.tb01375.x]
51. Moan IS, Rise J. Quitting smoking: applying an extended version of the theory of planned behavior to predict intention and behavior. J Applied Biobehav Res 2005;10(1):39-68. [doi: 10.1111/j.1751-9861.2005.tb00003.x]
52. Schifter DE, Ajzen I. Intention, perceived control, and weight loss: an application of the theory of planned behavior. J Pers Soc Psychol 1985 Sep;49(3):843-851. [Medline: 4045706]
53. Sheeran P, Orbell S. Implementation intentions and repeated behaviour: augmenting the predictive validity of the theory of planned behaviour. Eur J Soc Psychol 1999;29(2-3):349-369.
54. Cho H, Boster FJ. First and third person perceptions on anti-drug ads among adolescents. Commun Res 2008 Apr 01;35(2):169-189. [doi: 10.1177/0093650207313158]
55. Morrison L, Moss-Morris R, Michie S, Yardley L. Optimizing engagement with Internet-based health behaviour change interventions: comparison of self-assessment with and without tailored feedback using a mixed methods approach. Br J Health Psychol 2014 Nov;19(4):839-855 [FREE Full text] [doi: 10.1111/bjhp.12083] [Medline: 24308806]
56. Sarge MA, Knobloch-Westerwick S. Impacts of efficacy and exemplification in an online message about weight loss on weight management self-efficacy, satisfaction, and personal importance. J Health Commun 2013;18(7):827-844. [doi: 10.1080/10810730.2012.757392] [Medline: 23590282]
57. Kreuter MW, Chheda SG, Bull FC. How does physician advice influence patient behavior? Evidence for a priming effect. Arch Fam Med 2000 May;9(5):426-433. [Medline: 10810947]
58. Smit ES, Linn AJ, van Weert JCM. Taking online computer-tailoring forward: the potential of tailoring the message frame and delivery mode of online health behaviour change intervention. Eur Health Psychol 2015;17(1):25-31 [FREE Full text]
59. de Bruijn GJ, Out K, Rhodes RE. Testing the effects of message framing, kernel state, and exercise guideline adherence on exercise intentions and resolve. Br J Health Psychol 2014 Nov;19(4):871-885. [doi: 10.1111/bjhp.12086] [Medline: 24417878]
60. Gallagher KM, Updegraff JA. Health message framing effects on attitudes, intentions, and behavior: a meta-analytic review. Ann Behav Med 2012 Feb;43(1):101-116. [doi: 10.1007/s12160-011-9308-7] [Medline: 21993844]
61. Leone LA, Campbell MK, Allicock M, Pignone M. Colorectal cancer screening and physical activity promotion among obese women: an online evaluation of targeted messages. J Health Commun 2012;17(10):1187-1203 [FREE Full text] [doi: 10.1080/10810730.2012.665422] [Medline: 22775294]
62. Oenema A, Tan F, Brug J. Short-term efficacy of a web-based computer-tailored nutrition intervention: main effects and mediators. Ann Behav Med 2005 Feb;29(1):54-63. [doi: 10.1207/s15324796abm2901_8] [Medline: 15677301]
63. Rhodes RE, de Bruijn GJ. How big is the physical activity intention-behaviour gap? A meta-analysis using the action control framework. Br J Health Psychol 2013 May;18(2):296-309. [doi: 10.1111/bjhp.12032] [Medline: 23480428]
64. Armitage CJ, Norman P, Alganem S, Conner M. Expectations are more predictive of behavior than behavioral intentions: evidence from two prospective studies. Ann Behav Med 2015 Apr;49(2):239-246. [doi: 10.1007/s12160-014-9653-4] [Medline: 25623893]
65. Updegraff JA, Sherman DK, Luyster FS, Mann TL. The effects of message quality and congruency on perceptions of tailored health communications. J Exp Soc Psychol 2007;43(2):249-257 [FREE Full text] [doi: 10.1016/j.jesp.2006.01.007] [Medline: 18958299]

Abbreviations

CHART: Carolina Health Assessment and Resource Tool
HIT: Human Intelligence Task
MOST: multiphase optimization strategy
MTurk: Amazon Mechanical Turk

PA: physical activity
SCT: social cognitive theory
SDT: self-determination theory

Edited by G Eysenbach; submitted 28.02.17; peer-reviewed by JR Bautista, A Pfammatter, E Smit, R Paxton; comments to author 15.07.17; revised version received 08.12.17; accepted 16.12.17; published 01.03.18.

Please cite as:
Valle CG, Queen TL, Martin BA, Ribisl KM, Mayer DK, Tate DF
Optimizing Tailored Communications for Health Risk Assessment: A Randomized Factorial Experiment of the Effects of Expectancy Priming, Autonomy Support, and Exemplification
J Med Internet Res 2018;20(3):e63
URL: https://www.jmir.org/2018/3/e63/
doi:10.2196/jmir.7613
PMID:29496652

Original Paper

Web-Based Activity Within a Sexual Health Economy: Observational Study

Katy ME Turner[1], BSc (Hons), PhD; Adam K Zienkiewicz[2], PhD; Jonathan Syred[3,4], MSc; Katharine J Looker[5], PhD; Joia de Sa[6], MPH; Michael Brady[7], MBChB; Caroline Free[8], MBChB, PhD, MSc; Gillian Holdsworth[6], MBChB, MPH, MSc; Paula Baraitser[7,9], MBBS, BSc, MA, MD, FFPH, FFSRH

[1]Bristol Veterinary School, University of Bristol, Bristol, United Kingdom

[2]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom

[3]King's Center for Global Health and Health Partnerships, King's College London, London, United Kingdom

[4]School of Population Health & Environmental Sciences, King's College London, London, United Kingdom

[5]School of Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom

[6]SH:24, London, United Kingdom

[7]Department of Sexual Health and HIV, Kings College Hospital NHS Foundation Trust, London, United Kingdom

[8]Department of Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom

[9]King's Center for Global Health and Health Partnerships, School of Population Health & Environmental Sciences, King's College London, London, United Kingdom

Corresponding Author:
Katy ME Turner, BSc (Hons), PhD
Bristol Veterinary School
University of Bristol
Churchill Building
Langford Campus
Bristol, BS40 5DU
United Kingdom
Phone: 44 1173319120
Email: Katy.Turner@bristol.ac.uk

Abstract

Background: Regular testing for sexually transmitted infections (STIs) is important to maintain sexual health. Self-sampling kits ordered online and delivered in the post may increase access, convenience, and cost-effectiveness. Sexual health economies may target limited resources more effectively by signposting users toward Web-based or face-to-face services according to clinical need.

Objective: The aim of this paper was to investigate the impact of two interventions on testing activity across a whole sexual health economy: (1) the introduction of open access Web-based STI testing services and (2) a clinic policy of *triage and signpost* online where users without symptoms who attended clinics for STI testing were supported to access the Web-based service instead.

Methods: Data on attendances at all specialist public sexual health providers in an inner-London area were collated into a single database. Each record included information on user demographics, service type accessed, and clinical activity provided, including test results. Clinical activity was categorized as a simple STI test (could be done in a clinic or online), a complex visit (requiring face-to-face consultation), or other.

Results: Introduction of Web-based services increased total testing activity across the whole sexual health economy by 18.47% (from 36,373 to 43,091 in the same 6-month period—2014-2015 and 2015-2016), suggesting unmet need for testing in the area. Triage and signposting shifted activity out of the clinic onto the Web-based service, with simple STI testing in the clinic decreasing from 16.90% (920/5443) to 12.25% (511/4172) of total activity, $P<.001$, and complex activity in the clinic increasing from 69.15% (3764/5443) to 74.86% (3123/4172) of total activity, $P<.001$. This intervention created a new population of online users with different demographic and clinical profiles from those who use Web-based services spontaneously. Some triage and signposted users (29.62%, 375/1266) did not complete the Web-based testing process, suggesting the potential for missed diagnoses.

Conclusions: This evaluation shows that users can effectively be transitioned from face-to-face to Web-based services and that this introduces a new population to Web-based service use and changes the focus of clinic-based activity. Further development is underway to optimize the triage and signposting process to support test completion.

(J Med Internet Res 2018;20(3):e74) doi:10.2196/jmir.8101

KEYWORDS

sexually transmitted diseases; testing; internet; self-sampling

Introduction

Regular testing for sexually transmitted infections (STIs) with rapid treatment and partner notification are important strategies to improve and maintain sexual health [1]. Testing for chlamydia, gonorrhea, HIV, and syphilis is traditionally delivered within sexual health clinics, but Web-based testing is increasingly part of the sexual health economy [2-5]. This reflects the English National Health Service (NHS) strategy on digital care to meet expanding health care demand within limited resources [6,7].

Web-based STI testing services may provide self-tests where users both collect samples and read the results, or self-sampling where users collect samples that are sent to the laboratory for processing [8]. They offer 24-hour access to testing without the need to visit a clinic and may facilitate effective use of clinic services by shifting simple testing (testing with no other clinical activity required) online and freeing clinic capacity for complex care [9]. Although there is increasing availability of internet-based ordering of STI tests in developed countries, such services are highly heterogeneous: some are targeted to specific risk groups (based on age, ethnicity, or sexual orientation), test for a single infection [10], may be fully integrated with existing health services [11] or completely independent. A randomized controlled trial in France reported an increase in testing uptake (29.2% in the intervention group vs 8.7% in the control group, risk ratio: 3.37, 95% CI 3.05-3.74) [12]. However, outcomes were assessed using different measures in the intervention and control group, and there was low follow-up [12]. One US study found that internet-based testing could facilitate testing of high-risk individuals who were not accessing clinic-based services [13].

This paper evaluates the impact of Web-based self-sampling services within a sexual health economy within the London Boroughs of Lambeth and Southwark, an inner London area, with high rates of sexual ill health [14]. Unmet need was present before the introduction of an Web-based service, with 17,000 people having turned away from all sexual health services in the area annually (local clinic data) because of insufficient clinical capacity to meet demand. The Web-based service (SH:24 [15]) in this area provides free access to testing for chlamydia, gonorrhea, HIV, and syphilis for users older than 16 years with no restrictions based on gender or sexual orientation. Users complete an order form with self-sampling kits delivered home. Test kits are tailored to gender and sexuality. They include written information and link to a video that explains the self-sampling process. Participants can text or request a call-back for questions or concerns. Nonreturners are sent reminders via an SMS text message (short message service, SMS) and additional test kits if required. Results are sent by SMS text message except HIV reactive results that are delivered by telephone. At the time of this study, all those with positive results are referred into clinics for treatment and partner notification and managed according to national guidelines. Notification (by SMS text message) and management (treatment and partner notification) of patients identified through online self-sampling is the same as for asymptomatic patients tested in the clinic.

This paper documents STI testing before and after the introduction of online self-sampling services available to those aged 16 years or above and resident in the area. It documents the impact of a subsequent change in policy at one clinic—Camberwell Sexual Health Service, a large service providing both contraception and diagnosis and management of STIs. The change in policy was designed to facilitate the transition of simple STI testing from the clinic to the online service. This is an emerging strategy, developed in response to the online service. The new clinic policy introduced a triage process managed by clinical support workers who assess clinical need and redirect asymptomatic users requesting STI testing alone to order their tests via the online service using tablets available in the clinic. After online ordering, clinic staff prepare the self-sampling packs immediately for users to take away. Users then follow the online self-sampling process described above. The policy is designed to release capacity within the clinical service and use this for the management of complex clinical need. It is also intended to change future user behavior so that the online service becomes a first choice for future testing. The evaluation presented here analyses the impact of these service developments on sexual health activity across the whole sexual health system.

The aim of this observational study was to investigate the effect of the real life resource allocation decisions made within clinics following the establishment of an effective online service. This will inform policy makers and commissioners about the potential impact of changes to service capacity and delivery.

Methods

Data Sources and Preparation

Records of clinic visits for all sexual health attendances in the London Boroughs of Lambeth and Southwark were collated from January 1, 2014 to September 31, 2016 from all sexual health service providers: genitourinary medicine (GUM) clinics, integrated clinics, community sexual and reproductive health clinics, and the online service. This covered a baseline period where there were no major changes to sexual health service provision, followed by implementation of online STI testing,

and then changes in clinic practice resulting from availability of the new service (Figure 1).

All records were anonymized to remove identifiable information and harmonized to generate a complete dataset of individual level clinic attendances (one record, per person, per day). Each record includes demographic information: unique user identification number, gender, age at visit, site of visit, ethnicity, area of residence (lower super output area code), sexual orientation, and clinical information (first or follow-up visits and up to 12 sexual health, six reproductive health, and five contraception method *codes* for clinic activity). Individual level clinic attendance data were collated and summarized as simple STI test performed (chlamydia, gonorrhea, HIV, and syphilis) or complex service required. Area of residence was summarized as Lambeth, Southwark, adjacent boroughs, other London, or out of London. We considered the impact of the changes on Lambeth and Southwark residents only as access to the online service was restricted to this group. Clinic activity was coded using the GUM clinic activity dataset version 2 codes (Multimedia Appendix 1), assigned by clinicians during or after the consultation. For clinic users, we assumed that symptoms were present if microscopy was recorded as an activity. Online users were directly asked a question about presence of symptoms by the online registration system, with advice to go to a clinic if symptoms were present. This question was worded as follows:

> *Are you showing any symptoms of a sexually transmitted infections? If you have symptoms, we advise that you go to a clinic. You can find more information about your local clinic here. You can find out more information on symptoms of sexually transmitted infections here. (additional information screen appears on click).*

Exclusion Criteria

Records were excluded from analysis if there were no codes associated with the clinic visit, or if individuals were prisoners, or younger than 16 years, or 100 years and older.

Definition of Attendance Types and Positivity

Individual level clinic activity data were collated and summarized as "simple STI test" (chlamydia, gonorrhea, HIV, and syphilis) or "complex visit" (Multimedia Appendix 1). By definition, all Web activity was a simple STI test. We identified both attendances where a simple STI test was provided, as well as the subset of attendances where *only* a simple STI test was provided. A complex visit could be defined as an examination or physical intervention being recorded (eg, surgical, vaccination, or gynecological) or the patient was symptomatic.

The positivity was calculated as the number of positive diagnoses or total test records, both for all simple STI tests and for each of the four infections (gonorrhea, chlamydia, HIV, and syphilis) separately; further details of the calculation of positivity are given in Multimedia Appendix 1, equation 1.

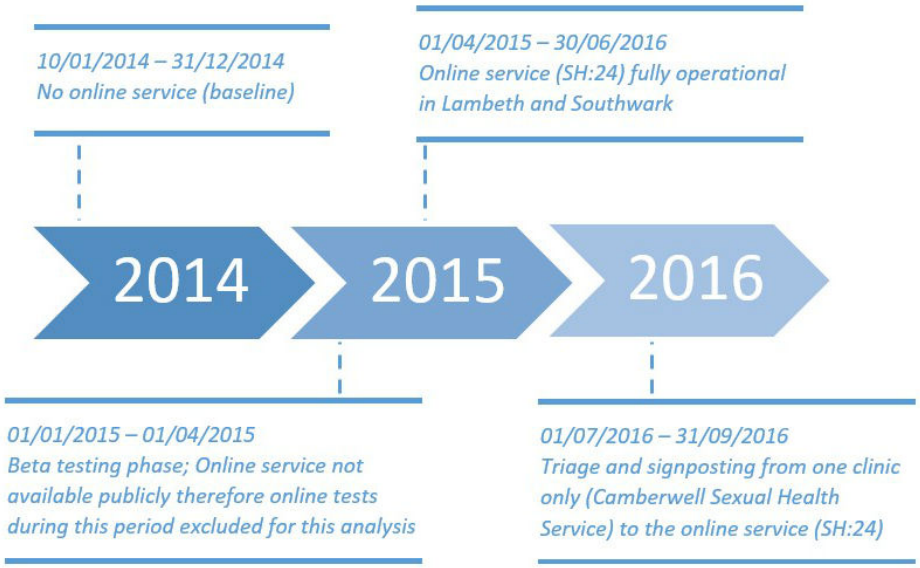
Ethics

Ethical approval was obtained from the NRES Committee North of Scotland—Grampian (Ref 15/NS/0031).

Data Analysis

The overall pattern of STI testing across Lambeth and Southwark across all sexual health providers from January 1, 2014 to September 30, 2016 was analyzed. Linear regression was used to test for trend in testing volume at Camberwell Sexual Health Centre up to June 30, 2016. The complexity of clinic activity provided to Lambeth and Southwark residents at Camberwell Sexual Health Centre in 2016 during quarter 2 (Q2) was compared with that during quarter 3 (Q3). The triage and signposting service was introduced at the start of Q3. Changes in pattern of clinic attendance between the quarters were analyzed using a chi-square test. Test completion and positivity were evaluated for those who used the triage and signposting pathway. Populations who used the triage and signposting pathway in Q3 were compared with those residents in the same area accessing the Web-based service without signposting or triage (spontaneous users).

Figure 1. Timeline of sexual health service changes in Southwark and Lambeth from 2014 to 2016.



Results

Patterns of Sexually Transmitted Infection Testing in Southwark and Lambeth Over Time

The overall testing capacity across the whole sexual health economy remained stable during the 2 years before the introduction of Web-based testing and is shown in Figure 2. The addition of online services was associated with an increase in total testing across the whole sexual health economy gradually from its introduction in Q2, 2015 to peak activity in Q3, 2016 (Figure 3). In a 6-month period before online testing (October 1, 2014 to March 31, 2015), there were a total of 36,673 STI tests performed within Lambeth and Southwark. In the same period in 2015-2016, this increased by 18.47% to 43,091 tests. Before the introduction of triage and signposting from January 1, 2014 to June 30, 2016, there was no significant change in the number of STIs tests each quarter over time in Camberwell Sexual Health clinic, test for trend *P*=.97 (Figure 3) .

We compared clinic activity at Camberwell Sexual Health Service, following the introduction of triage and signposting to direct asymptomatic users in the clinic to the Web-based service, implemented from July 1, 2016 (Table 1). The introduction of triage and signposting was associated with a decrease in total activity in the clinic of 23% between Q2 (6946) and Q3 (5362). The total number of visits by Lambeth and Southwark resident that included an STI test decreased from 3156/5443 (57.98% of visits) in Q2 to 2202/4172 (52.78% of visits) in Q3. During the same period, the proportion of simple STI tests without additional complex activity decreased from 16.90% (920/5443) to 12.25% (511/4172; chi-square, *P*<.001), and the proportion of complex service activity increased from 69.15% (3764/5443) to 74.86% (3123/4172; chi-square, *P*<.001; Figure 3).

Analysis of Sexually Transmitted Infection Testing Patterns in Quarter 3, 2016 Online and in Clinic

In the 3-month period after the introduction of the triage and signposting service, 2202 users obtained STI testing from Camberwell Sexual Health Service, 1266 users were signposted from this clinic after triage to the Web-based service, and 5362 used the Web-based service spontaneously. In addition, 175 people were triaged from another local sexual health service (Burrell St Clinic), who were excluded from subsequent analyses.

The introduction of the Web-based service and the triage and signposting service resulted in three different groups of service users (Table 2). For age, gender, and ethnicity, the clinic and the spontaneous online group showed differences in composition. The triage and signpost group was intermediate between the clinic and Web-based groups for age and ethnicity. For example, young people aged 16 to 19 years formed 8.89% (201/2261) of the clinic group, 4.55% (194/4262) of the spontaneous online group, and 6.2% (55/890) of the triage and signpost group.

The spontaneous Web-based group are most likely to be female: women formed 59.27% (1340/2261) of the clinic group, 64.43% (2746/4262) of the spontaneous Web-based group, and 46.7% (416/890) of triage and signpost group. Men who have sex with men formed a similar proportion of men using the clinic (30.4%, 251/825) and spontaneous Web-based (33.3%, 505/1516) but smaller proportion of triage and signposted (17.7%, 84/474), potentially reflecting higher probability of symptomatic infection or reported high risk behavior in the clinic population.

Both spontaneous Web-based service users and those registering at the clinic were encouraged to use clinic services if they had symptoms, and this is reflected in our data with 92.59% (3946/4262) of spontaneous Web-based users and 90.1% (802/890) of users who were triaged and signposted online classified as asymptomatic compared with 69.39% (1569/2261) of clinic users.

Figure 2. Number of simple sexually transmitted infections (STI) tests delivered across the whole sexual health economy, by service provider, by quarter, from quarter 1 (Q1) 2014 to quarter 3 (Q3) 2016 in Lambeth and Southwark, London.

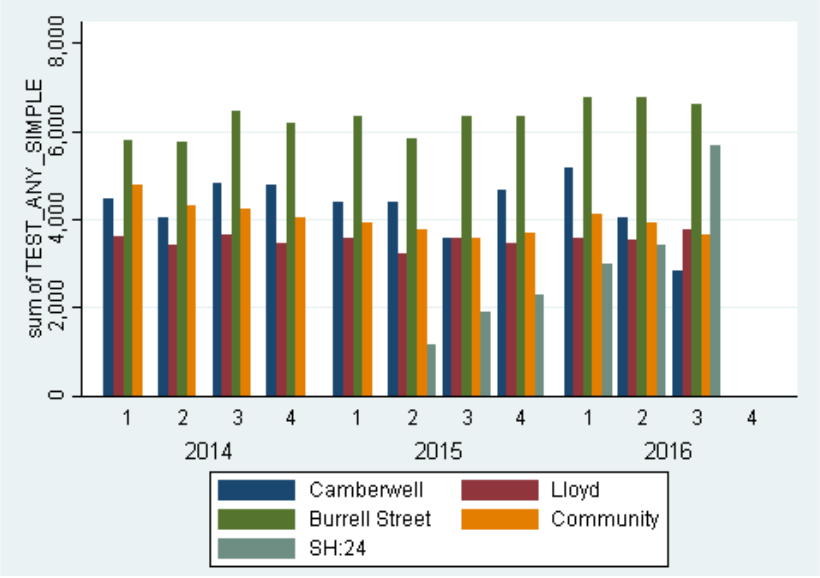


Figure 3. Number of simple sexually transmitted infections (STI) tests performed at Camberwell Sexual Health Centre and SH:24 by calendar year and quarter, quarter 1 (Q1) 2014 to quarter 3 (Q3) 2016.

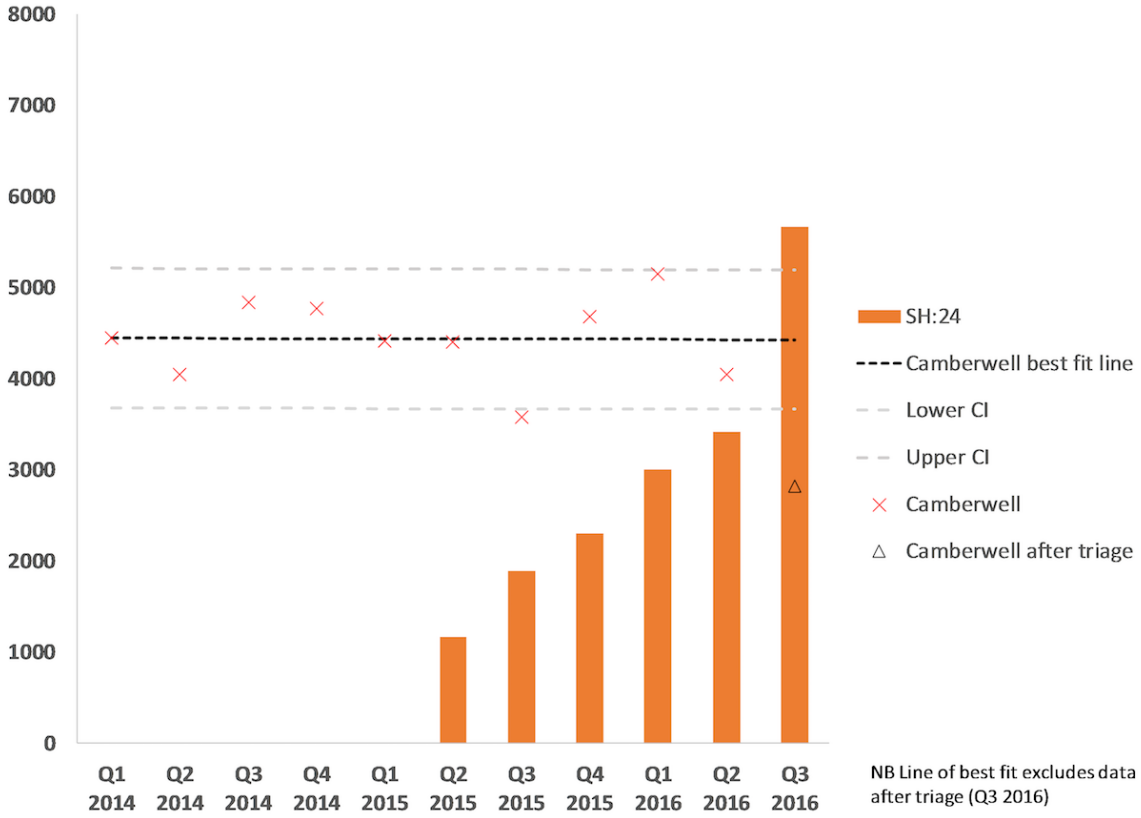


Table 1. Changing pattern of activity at Camberwell Sexual Health Service (quarter 2 [Q2] and quarter 3 [Q3] 2016) and Web-based testing.

Description	Camberwell		SH:24 Web-based service	
	Q2, N or n (%)	Q3, N or n (%)	Q2, N	Q3, N
Total records (1 per person per day)	7662	6245	4439	7717
Included records (age≥16 years, <100 years, and no prisoners)	7593	6188	4439	7716
Total visits with any code (all users)	6946	5362	4439	7716
Total visits (Lambeth and Southwark residents)	5443	4172	4436	7073
Total STI[a] tests (without complex intervention)	920 (16.90)	511 (12.25)	4436[b]	7073[b]
Total complex visits (with and without STI test)	3764 (69.15)	3123 (74.86)	0	0
Subtotal complex visits with STI test	2236	1691	0	0
Subtotal complex visit, no STI test	1528	1432	0	0
Other services used	759 (13.94)	538 (12.95)	0	0

[a]STI: sexually transmitted infection.

[b]Number of test kits sent out.

The test return rates within 6 weeks of test were higher for the spontaneous (70.51%, 3871/5632) than the "triage and signposting" groups (66.98%, 848/1266), which was significant (*P*=.01). A supplementary analysis comparing the demography of individuals who did and did not complete the testing process is shown in Multimedia Appendix 1. Women were less likely to complete tests, but otherwise, there were no significant differences between completers and noncompleters based on age, ethnicity, or sexual orientation, although the sample size is relatively small.

There were differences in positivity for any infection between the population that was seen in the clinic (7.70%, 174/2261), the *triage and signpost* users (6.4%, 57/890), and the *spontaneous* online users (4.58%, 195/4262). Patients diagnosed through Web-based testing were directed to clinics for management and partner notification in the same way as asymptomatic patients tested in the clinic. Any patients with HIV reactive tests are contacted to arrange confirmatory testing in the clinic.

Table 2. Characteristics of service users testing for sexually transmitted infections (STIs) via different pathways, July to September (quarter 3) 2016. A simple STI test includes chlamydia, gonorrhea, HIV, and syphilis.

Lambeth and Southwark residents (n=11,070)	Camberwell Sexual Health Clinic (n=4172)	SH:24 Web-based service (n=6898)[a]	
		Spontaneous online (n=5632)	Triage and signpost (n=1266)
Total visits, n	4172	5632	1266
Total tests ordered, n	2202	5632	1266
Return rate, n	N/A[b]		
≤2 weeks, n (%)		3186 (56.57)	775 (61.22)
>2 to ≤6 weeks, n (%)		785 (13.94)	73 (5.77)
>6 weeks, n (%)		293 (5.20)	43 (3.40)
Not returned, n (%)		1368 (24.29)	375 (29.62)
Total tests completed, n	2202	4262	890
Total tests completed plus STI diagnoses[c], n	2261	—	—
Age group, years, n (%)			
16-19	201 (8.89)	194 (4.55)	55 (6.2)
20-24	472 (20.88)	1282 (30.08)	205 (23.0)
25-29	508 (22.47)	1605 (37.66)	262 (29.4)
30-34	365 (16.14)	650 (15.25)	146 (16.4)
35+	715 (31.62)	531 (12.46)	222 (24.9)
Gender, n (%)			
Female	1340 (59.27)	2746 (64.43)	416 (46.7)
Male	921 (40.73)	1516 (35.57)	474 (53.3)
Men who have sex with men	251 (27.3)	505 (33.31)	94 (19.8)
Ethnicity, n (%)			
White	825 (36.49)	2850 (66.87)	461 (51.8)
Mixed	178 (7.87)	353 (8.28)	93 (10.4)
Asian	59 (2.61)	107 (2.51)	21 (2.4)
Black or black British	939 (41.53)	768 (18.02)	270 (30.3)
Other	208 (9.20)	105 (2.46)	32 (3.6)
Missing or prefer not to say	52 (2.30)	79 (1.85)	13 (1.5)
Symptoms, n (%)			
Asymptomatic	1569 (69.39)	3946 (92.59)	802 (90.1)
Symptomatic	692 (30.61)	316 (7.41)	88 (9.9)
Infection			
Positivity, any infection[d], n (%)	174 (7.70)	195 (4.58)	57 (6.4)
Chlamydia diagnoses, n	122	166	48
Gonorrhea diagnoses, n	49	23	4
Syphilis diagnoses, n	8	7	6
HIV diagnoses, n	2	4	2

[a]175 referrals from Burrell St excluded.

[b]N/A: not applicable.

[c]with no test during visit.

[d]Separate STI diagnoses do not add up to the total as some individuals were diagnosed with multiple infections. Positivity is only indicated for any infection, not shown for each separate infection as the number of tests completed was different for each infection.

Discussion

Principal Findings

The key findings from this evaluation are that availability of Web-based testing increased the total volume of STI testing and increased the proportion of clinic visits which utilized a *complex* service requiring face-to-face clinical evaluation. Once established as an effective method of testing, the availability of Web-based STI testing resulted in a change in clinic policy to actively signpost clinic attendees to use the Web-based service, reflecting high trust in the Web-based service, but also high demands on the clinic services. The majority (70.30%, 890/1266) of clinic users who were signposted to the Web-based service successfully completed their test. Almost a third (29.62%, 375/1266) of clinic attendees did not successfully complete an STI test.

This evaluation shows that users can effectively be transitioned from Web-based to face-to-face services and that this introduces a new population to Web-based service use and changes the focus of clinic-based activity.

Changing patterns of disease and rising user expectations are increasing demands on health services [7,16]. Supported self-management is one element of the response and can be provided through Web-based health services. Web-based self-management services work best when integrated with and supported by face-to-face care [17]. This creates interfaces between Web-based and terrestrial services that sustain effective functioning of these emerging hybrid systems.

Strengths and Limitations

Effective interaction with health services requires information and skills. This is acknowledged within sexual health services, with support for new service users such a young people to build capabilities for service access [18]. Self-management requires additional skills, and health services can support their development [19]. The triage and signposting intervention facilitated interaction with the Web-based service using strategies consistent with the literature on behavior change—making the transition easy, attractive, socially acceptable, and timely [20]. Tablets in the clinic and the test pack available immediately made the transition easy. Web-based services were attractive in that they avoided long waiting times in the clinic service. They were socially acceptable—with clinical staff promoting use and timely because they were offered at a time when users had identified a need for testing by visiting the clinic. By building capacity to self-manage STI testing, the service promotes a partnership approach to sexual health care delivered through the combined efforts of service providers and users [21,22]. However, nearly a third of those who were signposted to the Web-based service did not complete the testing process, and further service optimization is required to support shared decisions about clinic or Web-based service use that combine user preference, clinical guidance, evidence of efficient health service utilization, and that draw on a variety of user experiences. This work is in progress within this service using an agile approach with successive cycles of build-test-learn to optimize service design.

Across the whole sexual health economy, Web-based services increased total testing activity, suggesting a previously unmet need. During the study time frame, the testing volume in other service providers within the Southwark and Lambeth boroughs remained stable. Across England, there was a small increase in testing year on year (785,34—October 2014 to March 2015, increasing by 4% to 815,393—October 2015 to March 2016), which is in contrast to a 18.47% increase observed in Lambeth and Southwark. Before the introduction of triage and signposting, early adopters of Web-based STI testing were more likely to be women, aged 20 to 35 years, and of white ethnicity [23]. The triage and signposting intervention expanded the Web-based testing population compared with the group accessing via Web-based testing spontaneously. The residual population of clinic users were then more likely to have complex needs. Over a quarter of users who were signposted to the website (ordering a test with support in the clinic) did not complete the self-sampling and return a sample for testing. Women were less likely to complete the test in this group than men (Multimedia Appendix 1). A similar proportion of individuals using the Web-based service spontaneously also did not complete their tests. Noncompletion of tests for users in the triage and signposting group suggests possibility of missed diagnoses; however, it is not possible to ascertain whether these individuals were tested in other settings, for example, general practitioner or remained untested.

Triage and signposting changed patterns of unmet need in the sexual health economy studied. The continued increase in testing volume did not saturate during the study period, suggesting a continued unmet need in this population. The users who were previously turned away were advised to try another clinic or try the same clinic at a different time. The triage and signposting system offers these users a new option.

The intervention responds to two explicit objectives for the NHS: (1) to increase efficient use of resources and (2) to deliver user-centered care [7]. The evaluation suggests that it is partially successful on both counts. By focusing face-to-face clinical resource on complex need and shifting simple activity to supported self-management, it improves the efficiency of the sexual health economy. By building the capacity to use Web-based services, it offers an additional choice for STI testing that is potentially more convenient and more accessible [9,24].

Finally, this evaluation underlines the importance of research on the interfaces between Web-based and face-to-face services within the context of a whole sexual health economy. It suggests that users may be willing and able to move between Web-based services and clinic-based services with further research needed on how and why users transition between service modalities.

Implications

This paper reports on an innovative service evaluated in a timely way to inform service development. The use of routinely collected data collected offers consistent information on all attendances at all services in the area.

This paper evaluates a clinic-led service improvement in a complex and changing environment rather than a planned research intervention. Some important questions such as the

subsequent STI testing behavior of those who did or did not transition to Web-based services could not be answered with routinely collected data. We were unable to compare rates of treatment and partner notification with the routine data; however, positive patients were managed in the same way following diagnosis according to relevant clinical guidelines. Some assumptions such as the use of microscopy as a proxy for symptoms among clinic users are not substantiated. Additionally, some variables were self-reported in clinic data and could be missing where online users were required to select gender and sexual orientation to determine which test kits to send out.

Further research is required to investigate the long-term changes in service use behaviors, for example, the choice of service for subsequent STI testing, the reasons that those signposted online do not make the transition, improved strategies for triage and signposting so that those signposted are more likely to shift online, and the cost effectiveness of the shift in activity for sexual health economies.

Acknowledgments

Guy's and St Thomas' Charity funded this study. The authors would like to thank Robert Deluca (Kings College Hospital NHS Foundation Trust) and Guy's and St Thomas' NHS Foundation Trust for providing them with data. KJL and KMET would like to thank the National Institute for Health Research (NIHR) Health Protection Research Unit in Evaluation of Interventions at the University of Bristol for research support.

The views expressed are those of the authors and not necessarily those of the National Health Service, the NIHR, the Department of Health, or Public Health England.

Conflicts of Interest

PB, MB, and GH are directors of SH:24, a community-interest company providing Web-based sexual health services.

Multimedia Appendix 1

Further data on coding, additional analysis, and definition of positivity.

[PDF File (Adobe PDF File), 289KB - jmir_v20i3e74_app1.pdf]

References

1. Public Health England. Gov.uk. 2016. Sexually transmitted infections and chlamydia screening in England, 2015 URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/559993/hpr2216_stis_CRRCTD4.pdf [WebCite Cache ID 6qZzFeoCs]
2. Gaydos CA, Dwyer K, Barnes M, Rizzo-Price PA, Wood BJ, Flemming T, et al. Internet-based screening for Chlamydia trachomatis to reach non-clinic populations with mailed self-administered vaginal swabs. Sex Transm Dis 2006 Jul;33(7):451-457. [doi: 10.1097/01.olq.0000200497.14326.fb] [Medline: 16652069]
3. Estcourt C, Gibbs J, Sutcliffe L, Gkatzidou V, Tickle L, Hone K, et al. The eSexual Health Clinic system for management, prevention, and control of sexually transmitted infectionsxploratory studies in people testing for Chlamydia trachomatis. Lancet Public Health 2017;2(4):e183-e190 [FREE Full text] [Medline: 29253450]
4. Levine DK, Scott KC, Klausner JD. Online syphilis testing--confidential and convenient. Sex Transm Dis 2005 Feb;32(2):139-141. [Medline: 15668622]
5. Rietmeijer CA, Shamos SJ. HIV and sexually transmitted infection prevention online: current state and future prospects. Sex Res Soc Policy 2007 Jun;4(2):65-73. [doi: 10.1525/srsp.2007.4.2.65]
6. Health and Social Care Information Centre. NHS Digital. 2015. Information and technology for better care; our strategy for 2015-2020 URL: https://digital.nhs.uk/article/249/Our-Strategy [WebCite Cache ID 6qZzeQoWO]
7. NHS. NHS England. 2015. NHS Five Year Forward View URL: https://www.england.nhs.uk/wp-content/uploads/2014/10/5yfv-web.pdf [accessed 2017-05-19] [WebCite Cache ID 6qa2VGNhT]
8. Harding-Esch EM, Hollis E, Mohammed H, Saunders JM. Self-sampling and self-testing for STIs and HIV: the case for consistent nomenclature. Sex Transm Infect 2017 Mar;93(2):445-448. [doi: 10.1136/sextrans-2016-052841] [Medline: 27811311]
9. Baraitser P, Syred J, Spencer-Hughes V, Howroyd C, Free C, Holdsworth G. How online sexual health services could work; generating theory to support development. BMC Health Serv Res 2015 Dec 05;15:540 [FREE Full text] [doi: 10.1186/s12913-015-1200-x] [Medline: 26637295]
10. Hottes TS, Farrell J, Bondyra M, Haag D, Shoveller J, Gilbert M. Internet-based HIV and sexually transmitted infection testing in British Columbia, Canada: opinions and expectations of prospective clients. J Med Internet Res 2012;14(2):e41 [FREE Full text] [doi: 10.2196/jmir.1948] [Medline: 22394997]
11. Gilbert M, Haag D, Hottes TS, Bondyra M, Elliot E, Chabot C, et al. Get checked... where? The development of a comprehensive, integrated internet-based testing program for sexually transmitted and blood-borne infections in British

Columbia, Canada. JMIR Res Protoc 2016 Sep 20;5(3):e186 [FREE Full text] [doi: 10.2196/resprot.6293] [Medline: 27649716]
12. Kersaudy-Rahib D, Lydié N, Leroy C, March L, Bébéar C, Arwidson P, et al. Chlamyweb Study II: a randomised controlled trial (RCT) of an online offer of home-based Chlamydia trachomatis sampling in France. Sex Transm Infect 2017 May;93(3):188-195. [doi: 10.1136/sextrans-2015-052510] [Medline: 28377422]
13. Chai SJ, Aumakhan B, Barnes M, Jett-Goheen M, Quinn N, Agreda P, et al. Internet-based screening for sexually transmitted infections to reach nonclinic populations in the community: risk factors for infection in men. Sex Transm Dis 2010 Dec;37(12):756-763 [FREE Full text] [doi: 10.1097/OLQ.0b013e3181e3d771] [Medline: 20644498]
14. Public Health England. 2017. Sexual and reproductive health profiles URL: http://fingertips.phe.org.uk/profile/sexualhealth/data#page/0/gid/8000057/pat/6/par/E12000007/ati/102/are/E09000028%20Accessed%2028/03/2017 [WebCite Cache ID 6vJc1Qq4d]
15. SH24. URL: https://www.sh24.org.uk/ [accessed 2017-11-28] [WebCite Cache ID 6vJJCxZPe]
16. The Kings Fund. The future is now URL: https://www.kingsfund.org.uk/reports/thefutureisnow/ [accessed 2017-05-19] [WebCite Cache ID 6qa13FUvt]
17. Nationalvoices. 2014. Supporting self-management: summarising evidence from systematic reviews URL: http://www.nationalvoices.org.uk/sites/default/files/public/publications/supporting_self-management.pdf [WebCite Cache ID 6qa11Yax1]
18. Stone N, Ingham R. When and why do young people in the United Kingdom first use sexual health services? Perspect Sex Reprod Health 2003;35(3):114-120. [Medline: 12866784]
19. May CR, Eton DT, Boehmer K, Gallacher K, Hunt K, MacDonald S, et al. Rethinking the patient: using Burden of Treatment Theory to understand the changing dynamics of illness. BMC Health Serv Res 2014;14:281 [FREE Full text] [doi: 10.1186/1472-6963-14-281] [Medline: 24969758]
20. Hallsworth M, Snijders V, Burd H, Prestt J, Judah G, Huf S. Halpern D Applying behavioral insights simple ways to improve health outcomes. 2016 Presented at: World Innovation Summit for Health; 2016; Doha, Qatar p. 1-200 URL: https://www.imperial.ac.uk/media/imperial-college/institute-of-global-health-innovation/Behavioral_Insights_Report-(1).pdf
21. McWilliam CL. Patients, persons or partners? Involving those with chronic disease in their care. Chronic Illn 2009 Dec;5(4):277-292 [FREE Full text] [doi: 10.1177/1742395309349315] [Medline: 19933246]
22. Stewart M. Patient-Centered Medicine: Transforming the Clinical Method. Milton Keynes: Radcliffe Publishing; 2003:1857759818.
23. Barnard S, Turner K, Looker K, Bakolis I, Free C, Baraitser P. Equity of access to online sexually transmitted infection self-sampling services in lambeth and Southwark: an early view of the data. Sex Transm Infect 2016;92. [doi: 10.1136/sextrans-2016-052718.126]
24. Minichiello V, Rahman S, Dune T, Scott J, Dowsett G. E-health: potential benefits and challenges in providing and accessing sexual health services. BMC Public Health 2013;13:790 [FREE Full text] [doi: 10.1186/1471-2458-13-790] [Medline: 23987137]

Abbreviations

GUM: genitourinary medicine
NHS: National Health Service
STI: sexually transmitted infection
Q1: quarter 1
Q2: quarter 2
Q3: quarter 3

Edited by G Eysenbach; submitted 26.05.17; peer-reviewed by J Gibbs, V Minichiello; comments to author 24.08.17; revised version received 19.10.17; accepted 30.11.17; published 07.03.18.

Please cite as:
Turner KME, Zienkiewicz AK, Syred J, Looker KJ, de Sa J, Brady M, Free C, Holdsworth G, Baraitser P
Web-Based Activity Within a Sexual Health Economy: Observational Study
J Med Internet Res 2018;20(3):e74
URL: https://www.jmir.org/2018/3/e74/
doi:10.2196/jmir.8101
PMID:29514776

Original Paper

Expert Coaching in Weight Loss: Retrospective Analysis

Stefanie Lynn Painter[1*], RD, DHEd; Rezwan Ahmed[1*], PhD; Robert F Kushner[1], MD; James O Hill[1], PhD; Richard Lindquist[1], MD; Scott Brunning[1], MS; Amy Margulies[1], RD

Retrofit, Inc, Chicago, IL, United States

[*]these authors contributed equally

Corresponding Author:
Stefanie Lynn Painter, RD, DHEd
Retrofit, Inc
123 N. Wacker Drive
Suite 1250
Chicago, IL, 60606
United States
Phone: 1 18007745962
Email: stefanie@retrofitme.com

Abstract

Background: Providing coaches as part of a weight management program is a common practice to increase participant engagement and weight loss success. Understanding coach and participant interactions and how these interactions impact weight loss success needs to be further explored for coaching best practices.

Objective: The purpose of this study was to analyze the coach and participant interaction in a 6-month weight loss intervention administered by Retrofit, a personalized weight management and Web-based disease prevention solution. The study specifically examined the association between different methods of coach-participant interaction and weight loss and tried to understand the level of coaching impact on weight loss outcome.

Methods: A retrospective analysis was performed using 1432 participants enrolled from 2011 to 2016 in the Retrofit weight loss program. Participants were males and females aged 18 years or older with a baseline body mass index of ≥25 kg/m², who also provided at least one weight measurement beyond baseline. First, a detailed analysis of different coach-participant interaction was performed using both intent-to-treat and completer populations. Next, a multiple regression analysis was performed using all measures associated with coach-participant interactions involving expert coaching sessions, live weekly expert-led Web-based classes, and electronic messaging and feedback. Finally, 3 significant predictors (P<.001) were analyzed in depth to reveal the impact on weight loss outcome.

Results: Participants in the Retrofit weight loss program lost a mean 5.14% (SE 0.14) of their baseline weight, with 44% (SE 0.01) of participants losing at least 5% of their baseline weight. Multiple regression model (R^2=.158, P<.001) identified the following top 3 measures as significant predictors of weight loss at 6 months: expert coaching session attendance (P<.001), live weekly Web-based class attendance (P<.001), and food log feedback days per week (P<.001). Attending 80% of expert coaching sessions, attending 60% of live weekly Web-based classes, and receiving a minimum of 1 food log feedback day per week were associated with clinically significant weight loss.

Conclusions: Participant's one-on-one expert coaching session attendance, live weekly expert-led interactive Web-based class attendance, and the number of food log feedback days per week from expert coach were significant predictors of weight loss in a 6-month intervention.

(J Med Internet Res 2018;20(3):e92) doi:10.2196/jmir.9738

KEYWORDS

body mass index; coaching; feedback; obesity; overweight; weight loss; weight reduction program

Introduction

Worldwide, 1.9 billion adults are classified as being overweight or obese with the United States leading the globe [1,2]. This preventable disease is considered the driver of rising health care costs, and the annual direct and indirect health care costs have risen to $1.42 trillion [2].

In 2014, the direct medical costs of health conditions caused by overweight and obesity amounted to US $427.8 billion [2]. Indirect costs, such as absenteeism or loss of productivity due to disease, totaled US $988.8 billion [2]. With 70.7% of US adults being overweight or obese, employers spend an additional US $4000 more per year on an employee with obesity than on a healthy weight employee through costs related to health care, productivity, and job absenteeism [3-5]. According to the 2017 Employer Health Benefits Survey, 85% of employers provide health and wellness programs to prevent and manage chronic diseases [6]. Employer-sponsored weight management programs come in a variety of packages, including self-guided, group coaching, and individualized coaching related to activity, nutrition, and behavior change [7-11].

Weight management programs offering coaches to support participants have been shown to be more effective in participant engagement and weight loss success [7-9]. Females are more successful with weight loss programs that include direct and protocol-driven coaching around diet, physical activity, and engagement, whereas males tend to underuse coaches [12,13]. However, both males and females do benefit from coaches to increase engagement and weight loss success [9,12,13].

Offering education around behavior change and accountability for adherence of implementing information learned is one benefit of providing coaches with weight management programs. Face-to-face coaching sessions with weekly email contact from a coach was successful in helping participants lose at least 10% of initial body weight [14]. Alternatively, offering weekly email behavior coaching and monthly individualized coaching telephone calls has also shown to improve adherence to health-related strategies, decrease health risk factors, and improve weight loss [15-17]. In addition to individualized coaching, weekly behavioral change lessons, weekly individualized self-monitoring feedback, and an Web-based community group have also been shown to increase likelihood of achieving 5% weight loss in 6 months, 10% weight loss in 12 months, and maintenance of weight loss over 2 years [11,18,19].

Self-monitoring is important in achieving greater weight loss [20]. Coach-provided individualized feedback around self-monitoring increases consistency in both men and women [20]. Personalization proves to be more effective than automated emails providing general health information or tips specifically around nutrition and behavior [21-24].

The purpose of this study was to analyze the participant and coach interaction in a 6-month weight loss intervention administered by Retrofit (see Multimedia Appendix 1), a personalized weight management and Web-based disease prevention solution. The interactions were evaluated for their association with weight loss to determine the level of impact on predicting weight loss outcomes. Additionally, each type of interaction was evaluated independently to assess the association between the interaction and weight loss to determine best practices for expert coaches.

Methods

Study Design

A retrospective analysis was performed to assess the impact of expert coaching during a 6-month weight loss intervention using deidentified data from the Retrofit weight loss program. Various measures were designed to quantify coach-participant interactions involving one-on-one expert coaching sessions, live weekly expert-led interactive Web-based classes, food and exercise log feedback, and electronic messages. All measures were included in a multiple regression analysis to predict weight loss during the intervention. Finally, 3 statistically significant (P<.001) expert coaching measures were analyzed in depth to understand the impact on weight loss outcome at 6 months. Western Institutional Review Board granted exemption to the study as it is a retrospective analysis with no identifiable protected health information.

Participants

Participants included paying customers of the Retrofit program who enrolled through an employer-sponsored program. Employers of participants had selected Retrofit as a subsidized weight management program for employees as part of their employer health benefits package. Customers were considered as eligible participants if they were at least 18 years of age; had a starting body mass index (BMI) of ≥ 25 kg/m^2; had signed up for the program between September 27, 2011, and December 31, 2016; and had provided at least 1 weight measurement beyond baseline measurement. A participant was considered to have completed the program if he or she provided a weight measurement at the 6th month of his or her program. A total 1432 customers satisfied all inclusion criteria to be study participants, and 1045 of the participants completed the program. No customer was removed or eliminated from the population due to a lack of weight loss in the program.

Program

The Retrofit weight loss program was designed with a 6-month weight loss phase with the option to continue into a maintenance program called Retrofit Next. The program (Multimedia Appendix 2) includes one-on-one expert coaching, unlimited coach interactions through electronic messaging, lifestyle patterns assessment, and personalized coaching content and plan. Expert coaches perform weekly reviews of participants' plan and self-monitoring data to provide personalized feedback. Participants have access to an expert-moderated Web-based community and are encouraged to attend live weekly expert-led interactive Web-based classes regarding topics of exercise, nutrition, and mind-set. Digital tools, including a mobile app, Web-based dashboard, activity tracker, and Wi-Fi scale, are provided for tracking behaviors related to weight, food, mood, steps, and exercise.

As part of the Retrofit weight loss protocol, all participants are offered 7 one-on-one expert coaching sessions, including an initial 60-min session and 30-min follow-up sessions. Coaching sessions were conducted via Web-based video call or mobile phone. All coaching sessions include education around the Retrofit philosophy and weight loss guiding principles associated

with nutrition, mind-set, exercise, and daily activities. In addition, each coaching session was used for coach-participant collaboration on current and desired health-related behaviors, goal setting to create individualized plans and strategies, and to come to an agreement on how the expert coach will hold the participant accountable to agreed-upon plans and strategies.

Participants were encouraged to weigh in, wear their activity tracker, log all food and beverages consumed, and communicate daily with their expert coach and in the Web-based community. Retrofit protocol required expert coaches to review a participant's food and exercise logs, step data, weight data, and progress toward plan goals a minimum of 1 time per week to provide personalized feedback. If a participant initiated a coaching conversation, the expert coach was required to respond within 24 hours.

Retrofit expert coaches were employed professionals with a master's or doctorate-level education in dietetics or nutritional sciences, exercise physiology, nursing, health education, counseling, or psychology. Expert coaches were certified in Retrofit's weight loss protocol and have completed yearly recertification, if applicable.

Measures

Weight

Participants were provided a Wi-Fi-enabled scale that securely transmitted weight data over the Internet to a Retrofit central data server. Participants' weight data were collected through the use of the provided wireless scale (92% of recorded weights) or self-reported entry (8%). Self-reported entry was permissible if a participant had difficulty setting up his or her Wi-Fi scale. Baseline weight was considered as the first weight measurement received from the participant, which was designated as the recording for week 1. Percentage of baseline weight lost at 6 months was calculated and used as the primary outcome.

Expert Coaching Sessions

Participants were provided 7 one-on-one expert coaching sessions over the 6-month weight loss program. Percentage of coaching sessions attended at 6 months was calculated to quantify participant's engagement with their coach and used as one of the primary metrics to indicate coaching impact on participant outcome. A secondary metric was calculated to measure the total time a participant spent in coaching sessions.

Live Weekly Expert-Led Interactive Web-Based Classes

Participants were provided 26 weekly Web-based classes (1 class per week) where an expert coach conducted a live Web-based class on a predetermined topic. Percentage of classes attended at 6 months was calculated to quantify participants' interest in gaining in-depth knowledge on a healthy lifestyle and weight management practices. A secondary metric was calculated to capture the total time a participant spent in weekly Web-based classes.

Coach-Participant Conversations

The total number of coach-participant conversations was calculated by counting all electronic messages including coach-initiated conversations, coach responses to participant-initiated conversations, and coach feedback on food or exercise logs. The total number of coach-participant conversation days was calculated by including all days when an expert coach sent at least 1 electronic message. The average conversation length per week was calculated by counting the average of total length of all electronic messages (in characters) sent in a week.

To evaluate the impact of food log feedback on weight loss outcome, we calculated several measures to capture coach-initiated electronic feedback messages that include evaluation and guidance in response to participants' food logs. Total number of food log feedback counts all food log feedback provided by coach, which are defined as an expert coach comment written directly on a participant's individual food log or weekly diary of food log entries entered through digital tools provided. The total number of food log feedback days was calculated by counting all days with at least 1 food log–related feedback from the expert coach. The average food log feedback length per week was measured by averaging the total length of all feedback messages (in characters) provided in a week. Similar to food log feedback, 3 measures for exercise log feedback were also calculated.

Finally, 3 measures were defined to measure participant engagement with coach. Similar to expert coach–initiated electronic message measures, the total number of participant-initiated electronic messages, the total number of participant-initiated electronic message days, and the average participant-initiated electronic message lengths per week were calculated.

Statistical Analysis

All measures associated with coach-participant interactions involving expert coaching sessions, weekly Web-based classes, and electronic messaging and feedback were included in a multiple regression analysis to predict weight loss during the 6-month intervention. The least informative covariates were successively removed from the model in a stepwise elimination procedure based on the Akaike information criterion [25]. The regression model included only the main effects; interactions were beyond the scope of this analysis. In addition, this study focused on analyzing 3 statistically significant (P<.001) coaching interactions that were determined to be significant predictors in a weight loss model.

Data analyses were performed using R version 3.2.3 [26], which included dplyr 0.4.3, ggplot2 2.1.0, data.table 1.9.6, and leaps 2.9 packages. We also conducted t tests of equal variance on continuous variables at baseline and subsequent time points for 2 group comparisons. One-way analysis of variance (ANOVA) was utilized to determine mean differences for greater than 2 group comparisons. Subsequent Tukey tests were conducted to determine mean differences. Chi-square analyses were performed to determine differences among categorical variables when appropriate. For intent-to-treat (ITT) analyses, we used a last observation carried forward imputation approach. Alpha was set at .05 for all statistical tests to determine statistical significance.

Results

The reported results are based on the retrospective analysis evaluating the effect of various coach-participant interactions during the Retrofit 6-month weight loss intervention using both the ITT (N=1432) and the completer (n=1045 participants) populations. First, a detailed analysis on different coach-participant interaction measures is provided to understand both coach and participant behavior over a 6-month weight loss intervention. Second, a multiple regression model is presented to capture interaction measures that significantly impact participant outcome at 6 months, and finally, an in-depth analysis is provided for the top 3 significant measures.

Baseline Characteristics

Table 1 shows the demographic details at baseline for both ITT and completer populations. Although not clinically meaningful, the completers had higher average age compared with the overall population (45.73 vs 44.39, *P*=.001). Although there are differences in starting weight between completer and noncompleter groups, there are no differences in BMI at baseline between both populations. Furthermore, there are no differences in the male and female distribution among the ITT and completer groups (females: 61% vs 63%, *P*=.33).

Weight Change at 6 Months

For ITT population, the average weight loss at 6 months was 5.14% (SE 0.12), and 44% of the participants lost 5% or more of their baseline weight (see Table 2). For completers, the average weight loss at 6 months was 6.15% (SE 0.17), and 54% of the participants lost 5% or more of their baseline weight. For both ITT and completers, there were no significant differences between males and females in terms of weight loss percentage or the percentage losing 5% or more weight at 6 months.

Understanding Coach-Participant Interaction

The detailed quantitative analysis of the interaction between expert coach and participant is presented in Table 3. In general, completers had more interaction with coaches than the ITT population. The higher percentage of attendance or higher amount of interaction of the completers could be due to length of time actively participating in the weight loss program. Note that the average time in program for the noncompleters was about 3 months (mean 92.45 days, SE 2.20). In our analysis of the participant behavior below, we will focus on the ITT population.

Participants attended 75% of the one-on-one expert coaching sessions. Females attended higher percentage of coaching sessions than males (78.37% vs 70.72%, *P*<.001). Participants attended about 41% of the weekly Web-based classes. There is a gender difference observed in weekly Web-based class attendance as females attended significantly higher percentage of classes than males (51% vs 32%, *P*<.001). Consequently, females spent significantly higher amount of total time (638 min vs 405 min, *P*<.001) in classes learning about exercise, nutrition, and mind-set behaviors.

Table 1. Baseline demographics and outcome at 6 months.

Baseline demographics	Intent to treat (N=1432[a]), mean (SD)	Completers (n=1045[b]), mean (SD)	Noncompleters (n=387[c]), mean (SD)	*P* value[d]
Age, years	44.39 (10.31)	45.73 (10.10)	40.79 (10.00)	<.001
Starting weight, kg	104.76 (22.46)	103.95 (22.03)	106.94 (23.45)	.03
Starting body mass index, kg/m^2	35.88 (6.56)	35.82 (6.46)	36.03 (6.81)	.62

[a]869 female, 563 male.

[b]655 female, 390 male.

[c]214 female, 173 male.

[d]Completer vs noncompleter.

Table 2. Weight loss outcomes at 6 months.

Population	Intent to treat			Completers		
	n (%)	Weight loss percentage, mean (SE)	Lost 5% or more of baseline weight, mean (SE)	n (%)	Weight loss percentage, mean (SE)	Lost 5% or more of baseline weight, mean (SE)
Overall	1432 (100.00)	5.14 (0.14)	44 (0.01)	1045 (100.00)	6.15 (0.17)	54 (0.02)
Gender						
Female	869 (60.68)	5.19[a] (0.14)	44[b] (0.02)	655 (62.68)	6.00[c] (0.17)	52[d] (0.02)
Male	563 (39.32)	5.06[a] (0.14)	43[b] (0.01)	390 (37.32)	6.40[c] (0.18)	55[d] (0.03)

[a]For ITT, the weight loss difference between female and male is not significant (*P*=.66).

[b]For ITT, the difference between percentage of female and male losing 5% is not significant (*P*=.73).

[c]For completers, the weight loss difference between female and male is not significant (*P*=.27).

[d]For completers, the difference between percentage of female and male losing 5% is not significant (*P*=.38).

Table 3. Coach-participant interaction measures at 6 months.

Interactions	Intent to treat (N=1432), mean (SE)	Completers (n=1045), mean (SE)
Expert coaching sessions		
Percentage of coaching sessions attended	75.36 (0.72)	85.99 (0.61)
Total time spent in coaching sessions, min	188.34 (1.61)	211.32 (1.41)
Live weekly expert-led interactive Web-based classes		
Percentage of class attended	40.74 (0.83)	52.92 (0.92)
Total time spent in class, min	546.70 (10.52)	663.31 (11.81)
Coach-participant conversations		
Number of coach messages	158.91 (2.36)	180.00 (2.82)
Number of coach message days	75.16 (0.65)	82.36 (0.65)
Coach message length/week, characters	1458.34 (13.79)	1434.34 (20.94)
Number of food log feedback	74.91 (1.82)	89.27 (2.26)
Number of food log feedback days	31.89 (0.50)	37.01 (0.56)
Food log feedback length/week	409.29 (6.69)	410.05 (7.56)
Number of exercise log feedback	16.69 (0.32)	19.21 (0.38)
Number of exercise log feedback days	12.89 (0.23)	14.6 (0.26)
Exercise log feedback length/week	187.56 (3.89)	180.42 (4.22)
Number of participant messages	48.89 (1.27)	58.54 (1.60)
Number of participant message days	29.02 (0.64)	34.67 (0.77)
Participant message length/week, characters	399.29 (9.71)	433.12 (12.37)

Table 4. Multiple regression models identifying predictors of weight loss at 6 months. Multiple regression model summary: R2=.158; adjusted R2=.152, *P*<.001.

Models	Coefficients		
	β (SE)	*t* (degrees of freedom=997)	*P* value
Percentage of coaching sessions attendance	−1.05 (0.21)	−4.90	<.001
Percentage of weekly class attendance	−.76 (0.21)	−3.66	<.001
Number of food log feedback days	−.92 (0.26)	−3.50	<.001
Total number of coach message days	.89 (0.31)	2.83	.005
Coach message length per week	.54 (0.17)	3.14	.002
Number of participant messages	.95 (0.56)	1.68	.09
Number of participant message days	−1.56 (0.60)	−2.59	.01

Furthermore, coach-participant conversations were reviewed to assess the amount of interactions over the 6-month program. On an average, an expert coach reached out to his or her participant with responses, food/exercise log feedback, or general weight management guidelines approximately 75 days within the 6-month program (about 3 times a week). In general, participants who were more engaged in the program by initiating more conversations or logged more food/exercise logs received higher amount of communication from coaches. In addition, females received higher number of coach messages than males (170.24 vs 141.42, *P*<.001).

As reported in Table 3, almost half of the coach conversations were food log feedback (74.91 out of 158.91 messages). Females received significantly higher number of food log feedback than males (81.96 vs 64.04, *P*<.001). As females logged a higher number of food logs capturing their daily food intakes, coaches provided a higher amount of feedback. Participants either initiated conversation or responded to coach messages at least once a week (33.28 days) on average. Females sent higher number of messages than males (56.41 vs 37.28, *P*<.001 *)*.

Multiple Regression Model for Coach-Participant Interactions

A multiple regression model was built to predict weight change at 6 months by including all interaction measures related to coaching sessions, weekly Web-based classes, and coach-participant conversations. In the backward stepwise elimination multiple regression analysis, the final model

(R^2=.158, P<.001) included 7 coach-participant interaction measures, in which 6 of the measures were identified as statistically significant predictors: percentage of coaching sessions completed (β=−1.05, SE 0.21, P<.001), percentage of class attended (β=−.76, SE 0.21, P<.001), number of food log feedback days (β=−.92, SE 0.26, P<.001), total number of coach message days (β=.89, SE 0.31, P=.005), coach message length per week (β=.54, SE 0.17, P=.002), and number of participant message days (β=−1.56, SE 0.60, P=.01). The best regression model containing 7 coach-participant interaction measures is reported in Table 4.

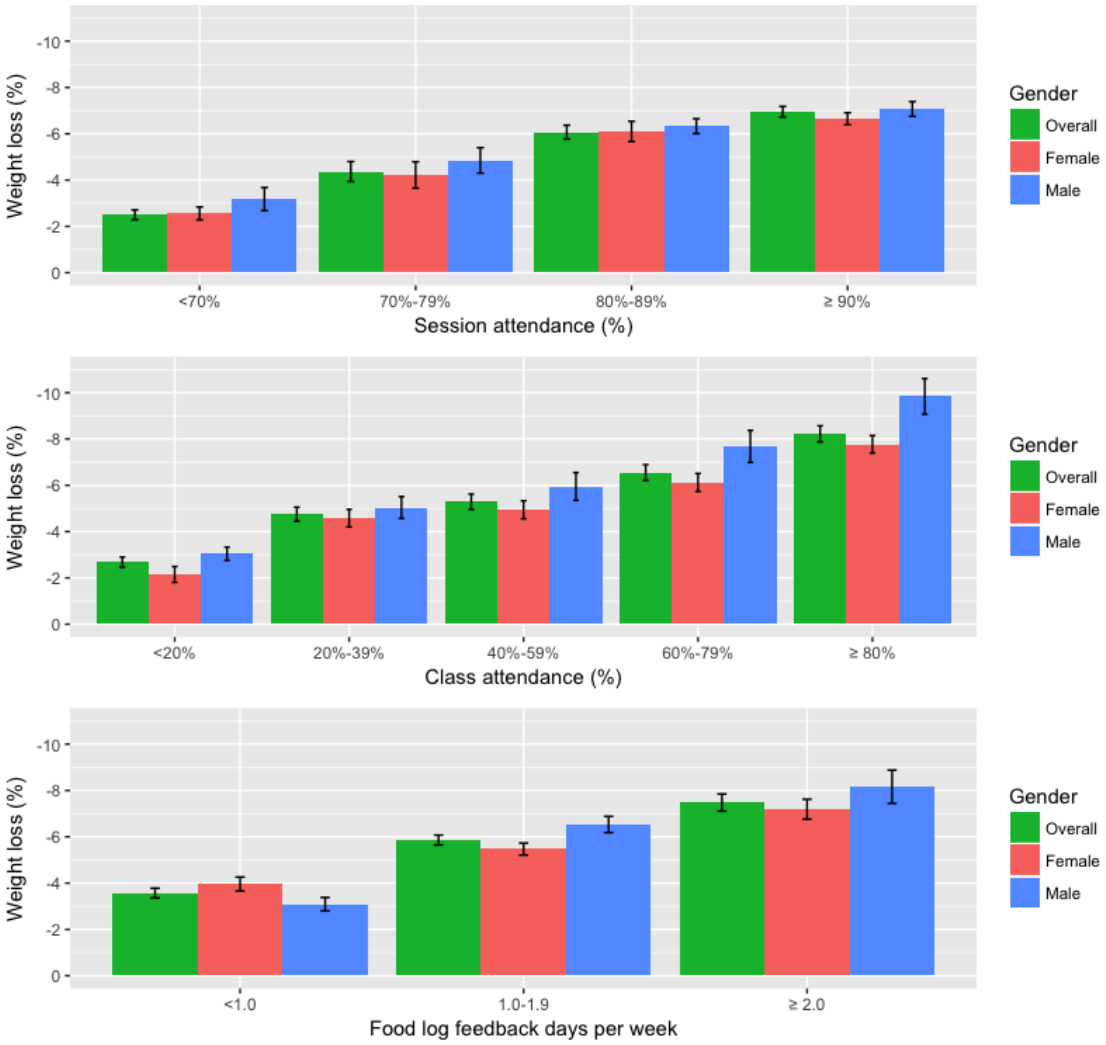
Significant Weight Loss Predictors: In-Depth Analysis

This section focuses on analyzing 3 of the predictors from the final regression model in Table 4, which have P<.001: percentage of coaching sessions completed, percentage of weekly classes completed, and number of food log feedback days. These analyses focus on quantifying different levels of coaching interaction and corresponding weight loss at 6 months to characterize the association with outcome. In addition, average coaching interactions were calculated for participants with different levels of weight loss at 6 months: lost ≥10% (264/1432, 18.44%), lost 5% to 10% (366/1432, 25.56%), and lost <5% (802/1432, 56.01%).

Expert Coaching Sessions

On the basis of the percentage of coaching session attendance data from the 6-month program, a higher percentage of coaching session attendance is significantly associated with a higher level of weight loss at 6 months. As shown in Figure 1, clinically significant weight loss (5%) was associated with at least 80% of coaching session attendance. The results of one-way ANOVA showed a significant difference of mean weight loss between different weigh-in levels (P<.001). A subsequent Tukey test confirmed the significant differences among the 80% to 90% and ≥90% attendance levels with the lower 2 levels (P<.001). Similar ANOVA tests were performed on male and female participants separately, and a significant difference in mean weight loss between different attendance levels was found (male: P<.001; female: P<.001).

Figure 1. Weight loss outcomes for different levels of coach-participant interaction.



Further analysis of coaching session attendance of participants with different levels of weight loss showed that a higher coaching session attendance was significantly associated with groups with higher levels of weight loss. Figure 2 shows a clear difference in coaching session attendance between loss <5% group and other 2 groups (P<.001). Both male and female participants separately showed a similar significant difference in coaching session attendance.

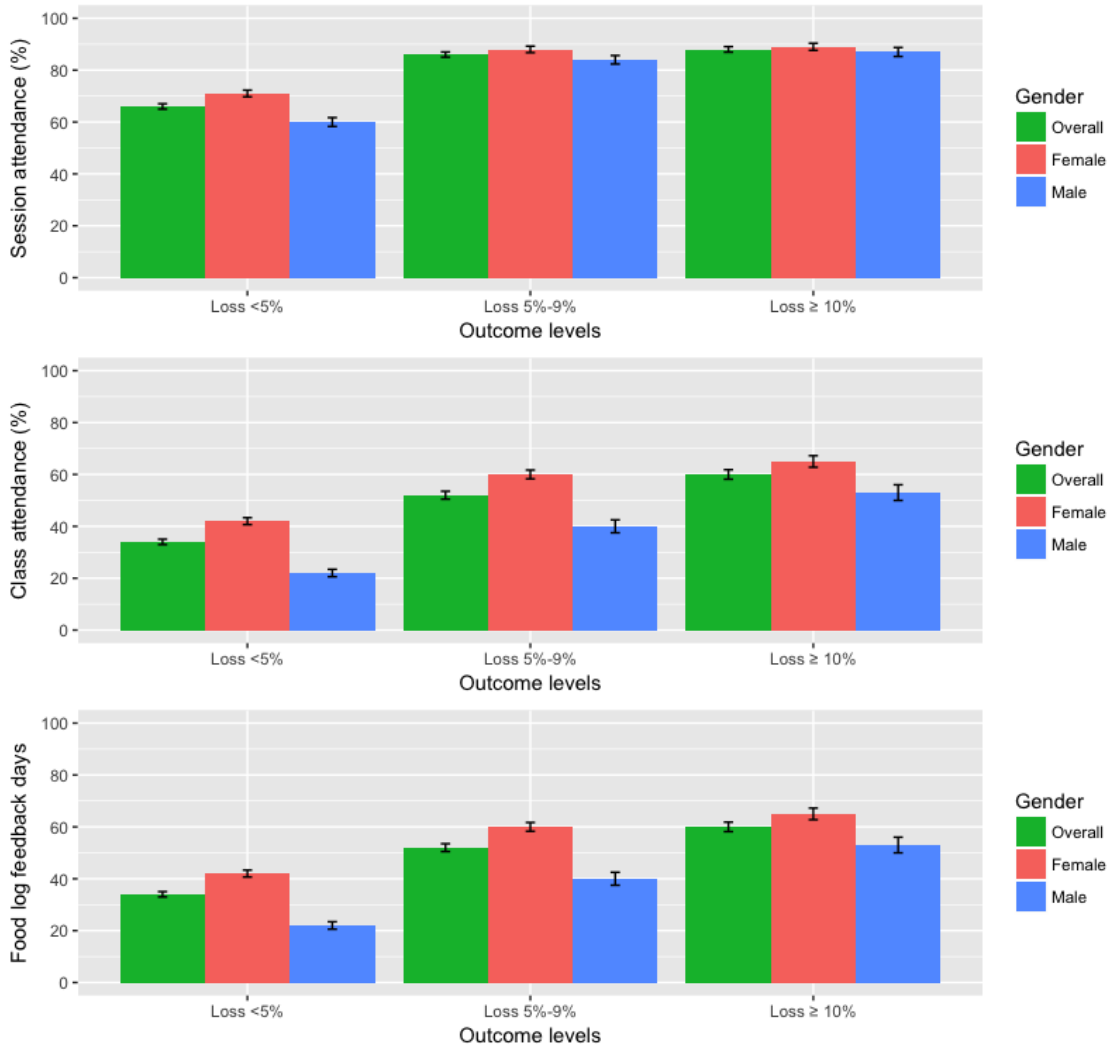
Live Weekly Expert-Led Interactive Web-Based Classes

As reported in Figure 1, the association between the percentage of weekly Web-based class attendance and weight loss at 6 months is linear where higher level of weight loss is significantly associated with higher percentage of class attendance. Clinically significant weight loss is associated with at least 60% of class attendance for overall and both male and females separately. One way ANOVA and a subsequent Tukey test confirmed significant mean differences in weight loss among 60% to 80% and ≥80% groups with the remaining levels of class attendance (P<.001). The analysis of percentage of class attendance of participants with different levels of weight loss showed that a higher class attendance was significantly associated with groups with higher levels of weight loss. Male and female participants separately showed similar significant differences in mean percentage of class attendance between different outcome levels (male: P<.001; female: P<.001).

Food Log Feedback Days

A higher number of food log feedback days per week is significantly associated with higher level of weight loss at 6 months. One way ANOVA test showed a significant mean difference in weight loss between difference in food log feedback levels (P<.001). A subsequent Tukey test confirmed significant mean differences between all levels of food log feedback days. Further analysis of food log feedback days of participants with different levels of weight loss showed that higher counts of food log feedback days were significantly associated with groups with higher levels of weight loss (P<.001).

Figure 2. Interaction levels of participants with different levels of outcome.



Discussion

Principal Findings

The results provide strong support for expert coaches in weight management programs. Participants had greater weight loss with a higher attendance of expert coaching sessions and live weekly expert-led interactive Web-based classes, as well as higher engagement with an expert coach through food log feedback. Completers also had greater interaction and attendance than ITT. In a multiple regression analysis, 6 of the 7 interaction measures were identified as statistically significant predictors of weight loss. In addition, an in-depth analysis of the top 3 significant predictors quantified the impact of coaching sessions completed, weekly Web-based class attendance, and days of receiving food log feedback on varying levels of weight loss. Overall, expert coaches were found to have a high impact on weight management.

Expert coaches provide guidance and accountability to increase participant engagement and weight loss success, which is supported by previous studies, including website, email, and/or mobile phone apps, as well as interventions using only phone calls for coaching [7-9,27]. However, the participant must be actively engaged in the program to receive benefit of the interactions. Quantifying the minimum and maximum level of engagement for significant weight loss can drive best practices for weight management expert coaches.

Although consistent self-monitoring is shown to have a predictive value for weight loss, the challenge is maintaining consistency among participants [20]. Findings support previous studies that personalized feedback and communication from expert coaches can produce greater engagement in self-monitoring activities when compared with tech-based interventions for self-monitoring without expert feedback [23,28,29]. We found that expert coaching sessions, live weekly expert-led classes, and food log feedback specifically increased interaction and have predictive weight loss values. On the basis of these results, it may be important to promote these coach-participant interactions together in an intervention or weight loss program.

Significant Predictors of Weight Loss

Expert Coaching Session Attendance

The percentage of coaching sessions completed was identified as a significant predictor of weight loss (P<.001). Attending 80% of the offered coaching sessions is associated with clinically significant weight loss of 5% or more. Others have shown that weekly to monthly coaching sessions are linked with 5% to 10% weight loss, improved adherence to health strategies, and decreased risk factors over a 6- to 12-month intervention [14-17]. Overall, female participants attended more coaching sessions than male participants, yet no significant difference was found in weight loss outcomes. Similar observations were reported in prior studies where male participants did not utilize expert coaches as frequently as female participants [12,13,15].

Live Weekly Expert-Led Interactive Web-based Classes

The percentage of weekly Web-based classes completed was identified as a significant predictor of weight loss (P<.001). Clinically significant weight loss of 5% is associated with at least 60% class attendance overall and between male and females separately. However, class attendance above 60% was associated with greater weight loss among all groups. Higher class attendance was linked to participants achieving 5% to 10% and >10% weight loss, yet male and female differed in class percentage attendance associated with levels of outcome. Males had a significantly lower attendance rate than females, which is historically common in weight loss interventions [12,13,15].

Food Log Feedback Days

The number of food log feedback days per week was identified as a significant predictor of weight loss (P<.001). Participants receiving food log feedback 1 to 2 days per week and ≥2 days per week were associated with clinically significant weight loss of 5% or greater. Additionally, participants in the 5% to 10% and >10% weight loss levels received more food log feedback days than those in the <5% weight loss level regardless of group. Food log feedback is directly dependent upon the participant's engagement in providing food logs for an expert coach to review. Females received a greater amount of feedback due to logging a higher number of food logs than males, which has been reported in earlier studies [20]. However, this finding is linked to the understanding that personalized feedback increases engagement and weight loss outcomes [22-24].

Strengths and Limitations

This study has several strengths, including the reporting of real-world weight loss outcomes and a focused analysis into expert coaches' role in a weight management program to determine which coach-participant interactions have a significant impact on participant success. Participants were existing participants of Retrofit and not recruited or incentivized to participate in the study. All participants who met the starting BMI, age, and weight criteria and provided at least 1 weight measurement beyond baseline were included as participants. No participant was removed from the population because of lack of success on the program, which is an uncommon research practice in the weight management field [30]. This study provides further insight on best practices of expert coaches in weight management interventions and programs. In addition, with the high population of male participants, gender comparisons were reported to create a greater understanding of interaction between male participants and coaches.

The study has limitations, which include the retrospective analysis study design that does not provide any causal inferences based on the critical observations. Coach-participant interaction was measured from a quantitative point of view. Also, the use of a real-world population does not reveal whether a participant was actively using any other weight management program outside of the Retrofit program components.

Future Research

Retrofit encourages all commercial weight loss programs to publish real-world research to enhance the understanding of coach-participant interactions in weight loss programs.

Reporting real-world data in relation to expert coaches allows commercial weight loss program to structure protocols for participant engagement and adherence to weight loss strategies. By fine-tuning interactions and by understanding how expert coaches are most effective, commercial weight loss programs will increase capability in overcoming the obesity crisis.

Recommended future research includes an analysis of specific strategies used by expert coaches and their impact on weight loss outcomes, as well as a qualitative analysis of the interactions between a coach and a participant, which may provide more insight into an expert coach's impact on participants. With the continued observation in this study and previous studies that male participants are less engaged than females, an analysis of strategies to increase male engagement and to understand whether increased engagement improves male weight loss outcomes is recommended. Additionally, further research is needed to analyze coaching impact on participants' self-monitoring behaviors to determine association between coach-participant interaction and the level of self-monitoring behaviors. Finally, expert coaches' impact beyond an initial 6-month intervention and the impact of each predictor of weight loss on weight maintenance would be a valuable future research study.

Conclusions

In conclusion, participants on the Retrofit weight loss program lost on average 5.14% (SE 0.14), and participants who completed the program lost on average 6.15% (SE 0.17) in 6 months. Over half of completers (54%) and 44% of all participants lost 5% or more of their baseline weight. Coach-participant interactions that include one-on-one expert coaching session attendance, live weekly expert-led interactive Web-based class attendance, and food log feedback days per week were shown to be significant predictors of weight change at 6 months. Specifically, attending 80% or more of offered expert coaching sessions, attending 60% or more of offered weekly Web-based classes, and receiving food log feedback one or more days per week from an expert coach increased participants' weight loss success.

Acknowledgments

Members of the Retrofit Advisory Board provided comments and professional insight around the data and results.

Conflicts of Interest

SP, RA, SB, and AM are employees of Retrofit, Inc, with equity in the company. JH, RK, and RL are active members of the Retrofit, Inc Advisory Board, with equity in the company.

Multimedia Appendix 1

Retrofit logo.

[JPG File, 64KB - jmir_v20i3e92_app1.jpg]

Multimedia Appendix 2

Features of the Retrofit Weight Loss program.

[JPG File, 2MB - jmir_v20i3e92_app2.jpg]

References

1. World Health Organization. Obesity and overweight URL: http://www.who.int/mediacentre/factsheets/fs311/en/ [WebCite Cache ID 6wjWaAWcr]
2. Waters H, DeVol R. Milken Institute. Weighing Down America: The Health and Economic Impact of Obesity URL: http://www.milkeninstitute.org/publications/view/833 [WebCite Cache ID 6vfARXaLA]
3. Gates DM, Succop P, Brehm BJ, Gillespie GL, Sommers BD. Obesity and presenteeism: the impact of body mass index on workplace productivity. J Occup Environ Med 2008 Jan;50(1):39-45. [doi: 10.1097/JOM.0b013e31815d8db2] [Medline: 18188080]
4. Van Nuys K, Globe D, Ng-Mak D, Cheung H, Sullivan J, Goldman D. The association between employee obesity and employer costs: evidence from a panel of U.S. employers. Am J Health Promot 2014;28(5):277-285. [doi: 10.4278/ajhp.120905-QUAN-428] [Medline: 24779722]
5. National Center for Health Statistics. Health, United States, 2016 URL: https://www.cdc.gov/nchs/hus/index.htm [WebCite Cache ID 6vdJUOX2i]
6. The Henry J. Kaiser Family Foundation. 2017 Employer Health Benefits Survey URL: https://www.kff.org/health-costs/report/2017-employer-health-benefits-survey/ [WebCite Cache ID 6vdJH45vu]
7. van Wier MF, Ariëns GA, Dekkers JC, Hendriksen IJ, Smid T, van Mechelen W. Phone and e-mail counselling are effective for weight management in an overweight working population: a randomized controlled trial. BMC Public Health 2009;9:6 [FREE Full text] [doi: 10.1186/1471-2458-9-6] [Medline: 19134171]

8. Vale MJ, Jelinek MV, Best JD, Santamaria JD. Coaching patients with coronary heart disease to achieve the target cholesterol: a method to bridge the gap between evidence-based medicine and the “real world”--randomized controlled trial. J Clin Epidemiol 2002 Mar;55(3):245-252. [Medline: 11864795]
9. Vale MJ, Jelinek MV, Best JD, Dart AM, Grigg LE, Hare DL, COACH Study Group. Coaching patients On Achieving Cardiovascular Health (COACH): a multicenter randomized trial in patients with coronary heart disease. Arch Intern Med 2003;163(22):2775-2783. [doi: 10.1001/archinte.163.22.2775] [Medline: 14662633]
10. Heinen L, Darling H. Addressing obesity in the workplace: the role of employers. Milbank Q 2009 Mar;87(1):101-122 [FREE Full text] [doi: 10.1111/j.1468-0009.2009.00549.x] [Medline: 19298417]
11. Xiao L, Yank V, Wilson SR, Lavori PW, Ma J. Two-year weight-loss maintenance in primary care-based Diabetes Prevention Program lifestyle interventions. Nutr Diabetes 2013 Jun 24;3:e76 [FREE Full text] [doi: 10.1038/nutd.2013.17] [Medline: 23797383]
12. Gabriele JM, Carpenter BD, Tate DF, Fisher EB. Directive and nondirective e-coach support for weight loss in overweight adults. Ann Behav Med 2011 Apr;41(2):252-263 [FREE Full text] [doi: 10.1007/s12160-010-9240-2] [Medline: 21108032]
13. Brace AM, Padilla HM, DeJoy DM, Wilson MG, Vandenberg RJ, Davis M. Applying RE-AIM to the evaluation of FUEL Your Life : a worksite translation of DPP. Health Promot Pract 2015 Jan;16(1):28-35. [doi: 10.1177/1524839914539329] [Medline: 24942749]
14. Leahey TM, Wing RR. A randomized controlled pilot study testing three types of health coaches for obesity treatment: professional, peer, and mentor. Obesity (Silver Spring) 2013 May;21(5):928-934 [FREE Full text] [doi: 10.1002/oby.20271] [Medline: 23784896]
15. Kirkman MS, Weinberger M, Landsman PB, Samsa GP, Shortliffe EA, Simel DL, et al. A telephone-delivered intervention for patients with NIDDM. Effect on coronary risk factors. Diabetes Care 1994 Aug;17(8):840-846. [Medline: 7956628]
16. Tate DF, Jackvony EH, Wing RR. Effects of Internet behavioral counseling on weight loss in adults at risk for type 2 diabetes: a randomized trial. J Am Med Assoc 2003 Apr 9;289(14):1833-1836. [doi: 10.1001/jama.289.14.1833] [Medline: 12684363]
17. Eakin EG, Lawler SP, Vandelanotte C, Owen N. Telephone interventions for physical activity and dietary behavior change: a systematic review. Am J Prev Med 2007 May;32(5):419-434. [doi: 10.1016/j.amepre.2007.01.004] [Medline: 17478269]
18. Tate DF, Wing RR, Winett RA. Using Internet technology to deliver a behavioral weight loss program. J Am Med Assoc 2001 Mar 7;285(9):1172-1177. [Medline: 11231746]
19. Carney D, Schultz S, Lim J, Walters W. Successful medical weight loss in a community setting. J Obes Weight Loss Ther 2015 Feb;5(1):pii: 248 [FREE Full text] [doi: 10.4172/2165-7904.1000248] [Medline: 26613067]
20. Painter SL, Ahmed R, Hill JO, Kushner RF, Lindquist R, Brunning S, et al. What matters in weight loss? An in-depth analysis of self-monitoring. J Med Internet Res 2017 May 12;19(5):e160 [FREE Full text] [doi: 10.2196/jmir.7457] [Medline: 28500022]
21. Hutchesson MJ, Tan CY, Morgan P, Callister R, Collins C. Enhancement of self-monitoring in a web-based weight loss program by extra individualized feedback and reminders: randomized trial. J Med Internet Res 2016 Apr 12;18(4):e82 [FREE Full text] [doi: 10.2196/jmir.4100] [Medline: 27072817]
22. Castro CM, Pruitt LA, Buman MP, King AC. Physical activity program delivery by professionals versus volunteers: the TEAM randomized trial. Health Psychol 2011 May;30(3):285-294 [FREE Full text] [doi: 10.1037/a0021980] [Medline: 21553972]
23. Tate DF, Jackvony EH, Wing RR. A randomized trial comparing human e-mail counseling, computer-automated tailored counseling, and no counseling in an Internet weight loss program. Arch Intern Med 2006;166(15):1620-1625. [doi: 10.1001/archinte.166.15.1620] [Medline: 16908795]
24. Sherrington A, Newham JJ, Bell R, Adamson A, McColl E, Araujo-Soares V. Systematic review and meta-analysis of internet-delivered interventions providing personalized feedback for weight loss in overweight and obese adults. Obes Rev 2016 Jun;17(6):541-551 [FREE Full text] [doi: 10.1111/obr.12396] [Medline: 26948257]
25. Akaike H. Bayes.acs.unt.edu. 1974. A New Look at the Statistical Model Identification URL: http://bayes.acs.unt.edu:8083/BayesContent/class/Jon/MiscDocs/Akaike_1974.pdf [accessed 2018-03-02] [WebCite Cache ID 6xcAM4IN2]
26. R-project. The R Project for Statistical Computing URL: https://www.r-project.org/ [accessed 2018-03-02] [WebCite Cache ID 6xbuvj8Jg]
27. Tao M, Rangarajan K, Paustian ML, Wasilevich EA, El Reda DK. Dialing in: effect of telephonic wellness coaching on weight loss. Am J Manag Care 2014 Feb 1;20(2):e35-e42 [FREE Full text] [Medline: 24738553]
28. Gold BC, Burke S, Pintauro S, Buzzell P, Harvey-Berino J. Weight loss on the web: a pilot study comparing a structured behavioral intervention to a commercial program. Obesity (Silver Spring) 2007 Jan;15(1):155-164. [doi: 10.1038/oby.2007.520] [Medline: 17228043]
29. Turk MW, Elci OU, Wang J, Sereika SM, Ewing LJ, Acharya SD, et al. Self-monitoring as a mediator of weight loss in the SMART randomized clinical trial. Int J Behav Med 2013 Dec;20(4):556-561 [FREE Full text] [doi: 10.1007/s12529-012-9259-9] [Medline: 22936524]
30. Gudzune KA, Bleich SN, Clark JM. Efficacy of commercial weight-loss programs. Ann Intern Med 2015 Sep 1;163(5):399. [doi: 10.7326/L15-5130-3] [Medline: 26322706]

Abbreviations

ANOVA: analysis of variance
BMI: body mass index
ITT: intent to treat

Edited by G Eysenbach; submitted 27.12.17; peer-reviewed by S Boucher, T Bose, A Haste; comments to author 17.01.18; revised version received 29.01.18; accepted 30.01.18; published 13.03.18.

Please cite as:
Painter SL, Ahmed R, Kushner RF, Hill JO, Lindquist R, Brunning S, Margulies A
Expert Coaching in Weight Loss: Retrospective Analysis
J Med Internet Res 2018;20(3):e92
URL: http://www.jmir.org/2018/3/e92/
doi:10.2196/jmir.9738
PMID:29535082

Original Paper

Treatment Activity, User Satisfaction, and Experienced Usability of Internet-Based Cognitive Behavioral Therapy for Adults With Depression and Anxiety After a Myocardial Infarction: Mixed-Methods Study

Emma Wallin[1,2], MSc (Psych); Fredrika Norlund[2], PhD (Psych); Erik Martin Gustaf Olsson[2], PhD (Psych); Gunilla Burell[3], PhD; Claes Held[2,4,5], MD, PhD; Tommy Carlsson[2], CCRN, RM, PhD

[1]Department of Psychology, Uppsala University, Uppsala, Sweden

[2]Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden

[3]Department of Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden

[4]Uppsala Clinical Research Center, Uppsala University, Uppsala, Sweden

[5]Department of Medical Science, Uppsala University, Uppsala, Sweden

Corresponding Author:
Emma Wallin, MSc (Psych)
Department of Psychology
Uppsala University
Box 1225
Uppsala, 75142
Sweden
Phone: 46 0706606118
Email: emma.wallin@psyk.uu.se

Abstract

Background: Knowledge about user experiences may lead to insights about how to improve treatment activity in Internet-based cognitive behavioral therapy (iCBT) to reduce symptoms of depression and anxiety among people with a somatic disease. There is a need for studies conducted alongside randomized trials, to explore treatment activity and user experiences related to such interventions, especially among people with older age who are recruited in routine care.

Objective: The aim of the study was to explore treatment activity, user satisfaction, and usability experiences among patients allocated to treatment in the U-CARE Heart study, a randomized clinical trial of an iCBT intervention for treatment of depression and anxiety following a recent myocardial infarction.

Methods: This was a mixed methods study where quantitative and qualitative approaches were used. Patients were recruited consecutively from 25 cardiac clinics in Sweden. The study included 117 patients allocated to 14 weeks of an iCBT intervention in the U-CARE Heart study. Quantitative data about treatment activity and therapist communication were collected through logged user patterns, which were analyzed with descriptive statistics. Qualitative data with regard to positive and negative experiences, and suggestions for improvements concerning the intervention, were collected through semistructured interviews with 21 patients in the treatment arm after follow-up. The interviews were analyzed with qualitative manifest content analysis.

Results: Treatment activity was low with regard to number of completed modules (mean 0.76, SD 0.93, range 0-5) and completed assignments (mean 3.09, SD 4.05, range 0-29). Most of the participants initiated the introduction module (113/117, 96.6%), and about half (63/117, 53.9%) of all participants completed the introductory module, but only 18 (15.4%, 18/117) continued to work with any of the remaining 10 modules, and each of the remaining modules was completed by 7 or less of the participants. On average, patients sent less than 2 internal messages to their therapist during the intervention (mean 1.42, SD 2.56, range 0-16). Interviews revealed different preferences with regard to the internet-based portal, the content of the treatment program, and the therapist communication. Aspects related to the personal situation and required skills included unpleasant emotions evoked by the intervention, lack of time, and technical difficulties.

Conclusions: Patients with a recent myocardial infarction and symptoms of depression and anxiety showed low treatment activity in this guided iCBT intervention with regard to completed modules, completed assignments, and internal messages sent to their therapist. The findings call attention to the need for researchers to carefully consider the preferences, personal situation,

and technical skills of the end users during the development of these interventions. The study indicates several challenges that need to be addressed to improve treatment activity, user satisfaction, and usability in internet-based interventions in this population.

(J Med Internet Res 2018;20(3):e87) doi:10.2196/jmir.9690

KEYWORDS

mental health; internet; cognitive behavioral therapy; computer-assisted therapy; myocardial infarction; attrition; adherence

Introduction

Background

Symptoms of depression and anxiety are common following a myocardial infarction [1,2]. These symptoms predict a worse somatic prognosis [3,4], and treatment and rehabilitation adherence [5,6], as well as poor quality of life [7]. Mental ehealth services such as guided internet-based cognitive behavioral therapy (iCBT) may improve access to acceptable, effective [8], and cost-effective interventions to reduce symptoms of depression and anxiety [8,9]. iCBT has also been found to improve psychological and physical functioning, as well as disease-related impact in chronic somatic conditions [10]. The use of eHealth solutions has received a growing interest as a suitable method in societies with limited health care resources and increasing numbers of aging individuals living with cardiac diseases [11]. There is preliminary evidence that iCBT may reduce symptoms of depression and anxiety among adults with high cardiovascular risk [12].

Typically, guided iCBT uses a written treatment material and internet-based synchronous or asynchronous communication with a therapist [13]. Compared with the traditional psychological care delivered face-to-face, internet-based interventions have several advantages such as reduced costs and increased user control and convenience [14]. Offering therapy that is more accessible with regard to place and time has the potential to make it easier for patients to fit a therapy into their daily life [15] and work according to their own preferred pace [16]. Moreover, internet-based interventions may be a way to reach people who feel embarrassed when talking to a care provider about their symptoms [14]. However, preliminary evidence suggests that people of higher age, which is associated with myocardial infarction [17], may experience more technical problems using internet-based interventions [18]. It has also been suggested that participants recruited in routine care may have less favorable views of internet-based interventions [19] and that those recruited through a consecutive clinical procedure are less motivated to engage in internet-based interventions compared with patients recruited through self-referral [20].

Treatment acceptability may be defined as the extent intended users perceive a given intervention as reasonable, justified, fair, and palatable [21]. Studies investigating iCBT have reported issues with indicators of treatment acceptability [22], including low expectations of its helpfulness and credibility [23,24], low take-up rates [25], high dropout rates [20,26], and poor adherence [27]. Treatment activity is an important aspect for internet-based interventions that aim to treat depression and anxiety disorders, as number of completed modules correlate with outcome [28]. There is an articulated need for qualitative studies conducted alongside quantitative trials, which investigate determinants of treatment acceptability [25]. Studies that explore experiences of taking part in internet-based interventions may lead to valuable insights on how to offer more effective treatments [29,30]. As little is known about the acceptability of iCBT interventions among patients with a recent myocardial infarction, there is a need for explorative studies to investigate treatment activity and experiences among such intended end users.

Objectives

The overall aim of this study was to explore treatment activity, user satisfaction, and usability experiences among patients allocated to treatment in the U-CARE Heart study, a randomized clinical trial of an iCBT intervention for treatment of depression and anxiety following a recent myocardial infarction (unpublished data, 2018; [31]). The following 2 research questions were addressed:

1. What was the treatment activity with regard to completed modules, completed assignments, and therapist communication initiated by the participants?
2. What positive and negative experiences of the intervention, as well as suggestions for improvement, did the participants describe?

Methods

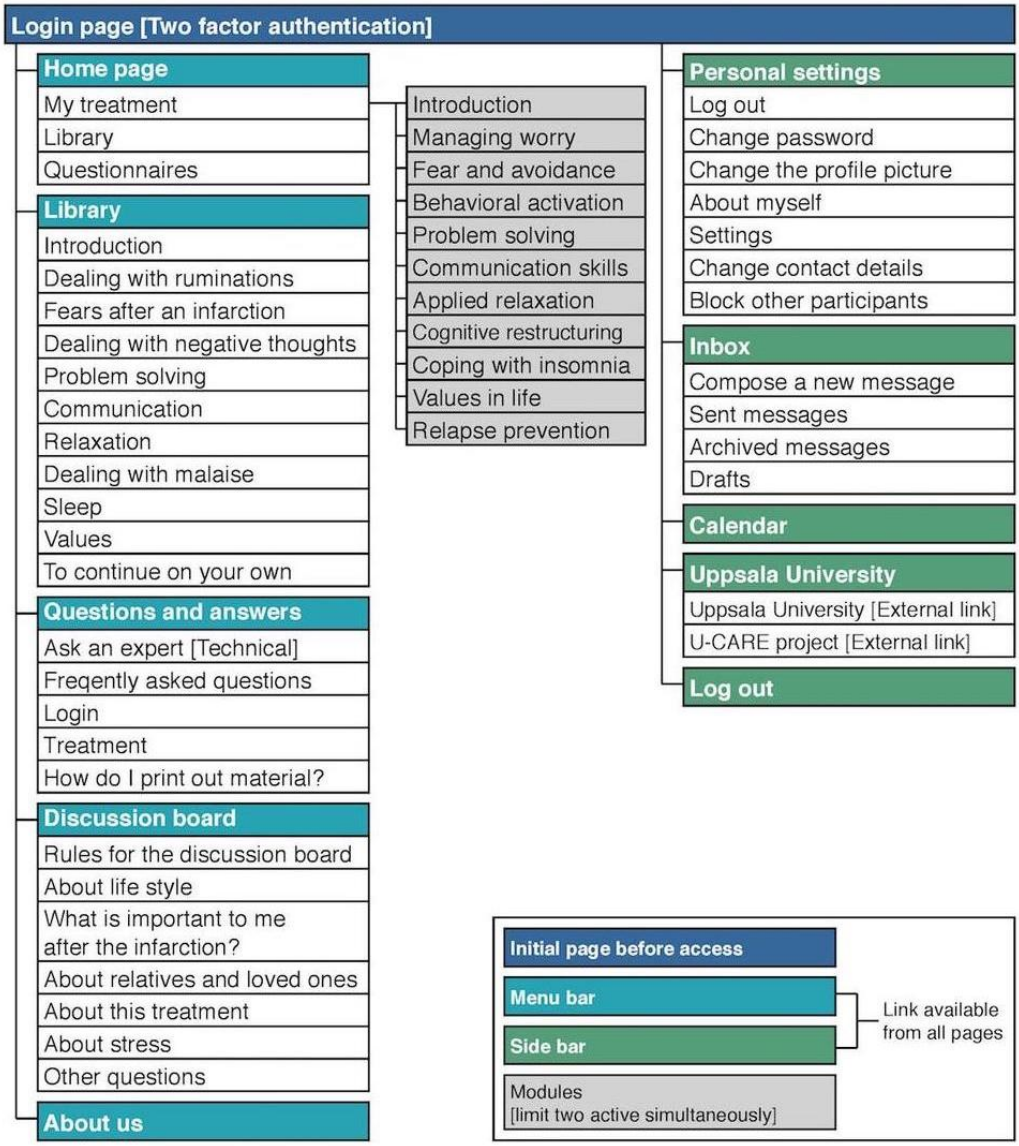
Study Design

This study was conducted alongside the U-CARE Heart study. The results from the randomized controlled trial (RCT) indicate no differences between the groups in symptoms of depression and anxiety after intervention [31]. This study is a descriptive mixed methods study, with quantitative and qualitative approaches. The regional ethics committee in Uppsala approved the study protocol (2011/217). The RCT was preregistered at ClinicalTrials.gov, Identifier: NCT01504191 December 2011.

The Internet-Based Cognitive Behavioral Therapy Intervention

The U-CARE Heart study used an internet-based portal to deliver an iCBT intervention tailored for patients with a recent myocardial infarction. A two-factor authentication solution with a password and numerical short message service (SMS) verification was required to log on to the portal. The design of the portal included a side bar and a menu bar, accessible from all pages. A short presentation and pictures of the therapist who worked in the program was provided in the “About us” section. Figure 1 presents a sitemap of the portal. Multimedia Appendix 1 presents a screenshot of the portal.

Figure 1. Sitemap of the internet-based portal.



The treatment program consisted of 11 modules. Each module consisted of 2 to 4 steps. Each step contained 1 or 2 assignments, such as self-monitoring or registration of skills training. The treatment material consisted of PDF files with psychoeducation. The average word count per module was 6739.91 (SD 2786.79). Participants were encouraged to work with one step per week during the 14-week treatment period. The first introductory module was mandatory and oriented the user to the portal and the treatment program through an instructional video, psychoeducation about CBT, and common reactions post myocardial infarction. Thereafter, participants were invited to read a short description of the available modules, before choosing which modules to work with. Participants were limited to work with 2 active modules simultaneously. Supplementary material and video clips of interviews conducted with patients about their experience of depression and anxiety after a myocardial infarction were available throughout the course of treatment in an additional module called the Library. Participants also had access to a discussion board where they could communicate with other participants.

Each patient was assigned 1 of 3 therapists, who could be contacted any time. Therapists provided asynchronous written feedback on assignment via an internal message function within 24 hours. After completing all steps in a module, approval from a therapist was needed to activate a new module. Participants inactive for more than 1 week were reminded to stay active via phone calls. Participants unable to be reached were reminded by prompts sent via SMS.

Clinical psychologists and experts in IT solutions developed a preliminary version of the intervention. This version was evaluated through face-to-face think-aloud sessions [32], through consultations [33] with 6 test users with experience from emotional distress after a myocardial infarction, 2 stress management groups in regular cardiac rehabilitation, and two of the cardiac nurses involved in recruitment of patients at cardiac clinics. A description of the results of these consultations is presented in Multimedia Appendix 2. The introductory module was slightly modified (shortened) after a pilot trial [34]. During the study, the portal was adapted for handheld devices after 63 participants had been randomized.

Recruitment

Randomized Controlled Trial

Patients were consecutively recruited from 25 cardiac clinics in Sweden. To be eligible, patients needed to: (1) be younger than 75 years, (2) have a medical history of a recent myocardial infarction less than 3 months prior, and (3) report a score >7 on either the depression or anxiety subscale in the Hospital Anxiety and Depression Scale (HADS) [35]. Potential participants were excluded if they: (1) had a life expectancy of less than 1 year, (2) were scheduled for bypass surgery, (3) were unable or unwilling to use a computer or mobile phone, (4) were unable to read or write in Swedish, (5) had an anticipated poor compliance to iCBT (eg, alcohol abuse), (6) had severe self-reported depression (total score >34) or risk of suicide (item 9 >3) on the Montgomery Asberg Depression Rating Scale Short form (MADRS-S) [36], or (7) participated in another ongoing trial with a behavioral intervention. The HADS and MADRS-S were administered via the internet. In total, 3928 persons were assessed for eligibility. Of these, 117 were allocated through randomization to the iCBT intervention (Figure 2).

Follow-Up Telephone Interviews

Participants (n=69) allocated to the treatment arm, between June 2015 and October 2016, were eligible to participate in a follow-up telephone interview (Figure 2). Participants were excluded if they had not filled out postintervention questionnaires in the RCT (n=4), or terminated treatment prematurely (n=3). Additionally, 6 participants were not approached because of administrative reasons. Of the approached participants (n=56), 13 declined participation, and 20 could not be reached or did not return the consent form. This resulted in 23 interviews. However, 2 interviews were excluded due to poor sound quality of the recordings, resulting in 21 interviews included in the qualitative analysis.

Sample Characteristics

The majority of the participants in the randomized trial were males, employed, living in a relationship, born in Sweden, had no children in their household, and did not receive any current counseling (Table 1). The baseline sample characteristics for participants in the randomized trial were represented in the sample of participants included in the follow-up telephone interviews with 2 exceptions. None of the participants who were unemployed or on sick leave were included in the follow-up interviews. Furthermore, a higher proportion of the participants included in the interviews had studied at a university compared with participants not included in the interview (P=.03).

Data Collection

Treatment Activity in the Randomized Controlled Trial

User activity was automatically registered through the U-CARE internet portal. Number of completed modules and assignments, and internal messages sent from patients to therapists, were used as quantitative measures of treatment activity.

Follow-Up Telephone Interviews

The fourth author (GB) conducted individual telephone interviews with the aid of a semistructured interview guide (Multimedia Appendix 3). Participants were informed that the purpose of the interviews was to evaluate their experiences of the intervention. Probes were asked to explore experiences and preferences. A preliminary interview guide was developed by three of the authors (EW, FN, and GB) and tested by interviewing 2 participants allocated to the intervention. These interviews were later included in the analysis. The phrasing of some of the questions in the guide was revised after these interviews. The interviews were audio-recorded, transcribed verbatim, and lasted between 22 and 66 min.

Data Analysis

Treatment Activity in the Randomized Controlled Trial

Quantitative data regarding number of completed modules, assignments, and therapist communication initiated by participants were analyzed with descriptive statistics using R version 3.2.2 (R Foundation for Statistical Computing).

Telephone Interviews

The interviews were analyzed with inductive qualitative manifest content analysis, inspired by the outline presented by Graneheim and Lundman [37]. Interviews were transcribed verbatim by a professional transcribing agency. Two authors (EW and TC) were responsible for the analysis. Initially, the interview transcripts were read multiple times to obtain an overall perspective of the content. Meaning units were identified, defined as words, sentences, or paragraphs of a single message or context that corresponded to positive experiences, negative experiences, or suggestions for improvement. These meaning units were condensed, so that unnecessary words were removed. Thereafter, the condensed meaning units were labeled with a code that represented the core content and context of the meaning unit. Codes were sorted into categories and subcategories of the manifest content, defined as collections of codes that shared a commonality with regard to the visible content, identified with as little interpretation as possible. Initially, both authors worked independently with 2 interview transcripts and discussed the identified meaning units, condensed meaning units, codes, and preliminary categories. No impactful differences were observed. Thus, the first author (EW) identified meaning units, condensed the identified meaning units, and labeled these with a code for the remaining transcripts. Repeated face-to-face discussions were held between the authors EW and TC, with the purpose of scrutinizing the findings from the perspectives of the last author (TC), who had no previous experience of the U-CARE Heart intervention. Codes were sorted into subcategories and categories with the aid of NVivo version 11.3.2 (QRS International Pty Ltd., Australia). Multimedia Appendix 4 presents examples of the steps in the qualitative analysis, and Multimedia Appendix 5 presents backgrounds of researchers involved in qualitative data collection and analysis.

Figure 2. Recruitment of participants. HADS-A: Hospital Anxiety and Depression Scale-Anxiety; HADS-D: Hospital Anxiety and Depression Scale-Depression.

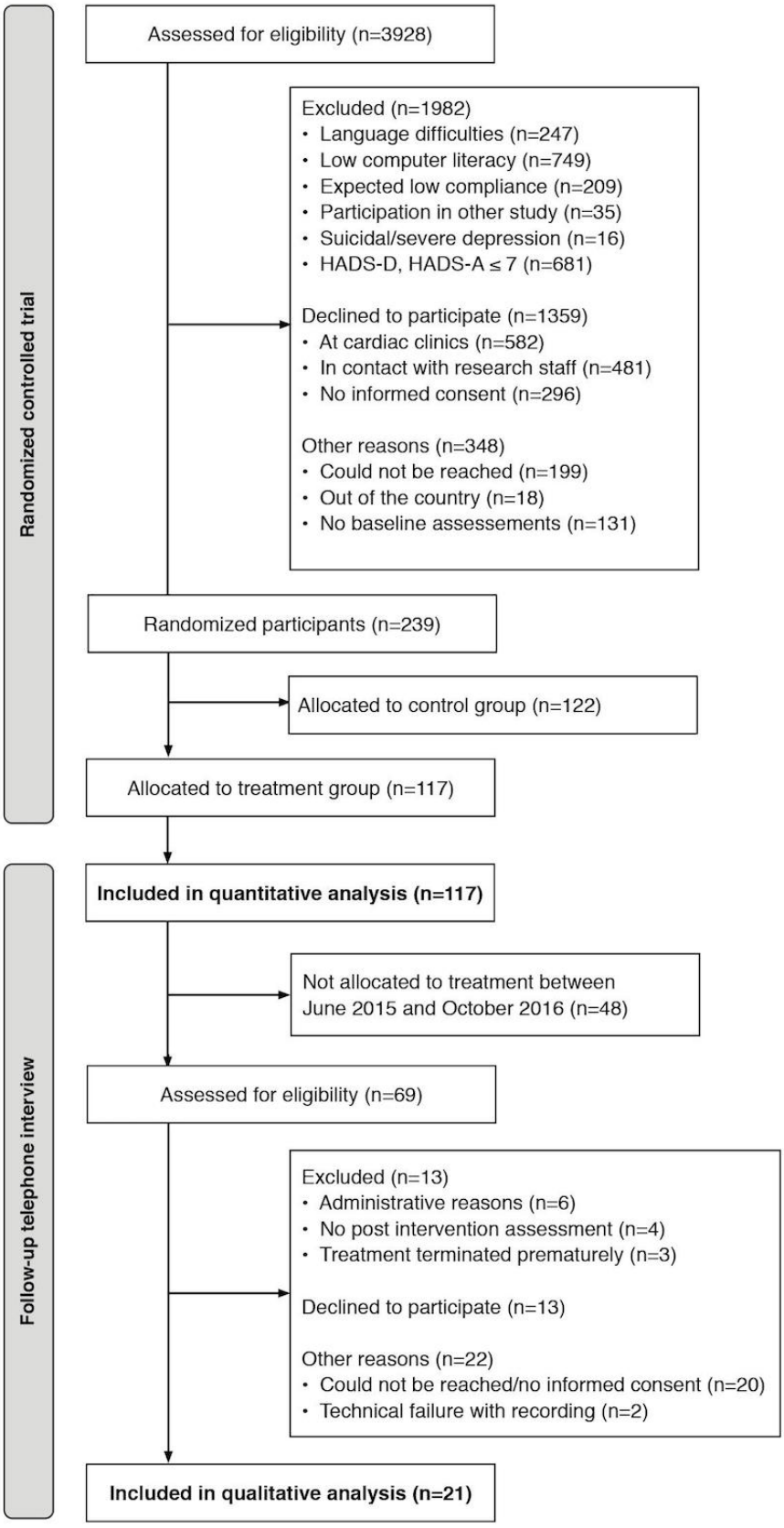


Table 1. Baseline demographic and clinical characteristics of participants. Between-group comparisons are conducted between participants not interviewed and interviewed. Categorical data is analyzed with Fisher exact test and continuous data is analyzed with Welsh *t* test.

Characteristics	Allocated to intervention (n=117)	Not interviewed (n=96)	Interviewed (n=21)	*P* value
Demographic characteristics				
Age in years, mean (SD)	58.37 (8.98)	58.68 (8.67)	56.95 (10.38)	.48
Sex, n (%)				
Female	44 (37.6)	37 (39)	7 (33)	.80
Male	73 (62.4)	59 (62)	14 (68)	
Occupation, n (%)				
Employed	78 (66.7)	62 (65)	16 (76)	.44[a]
Unemployed	4 (3.4)	4 (4)	0 (0)	
Retired	33 (28.2)	28 (29)	5 (24)	
Sick leave	2 (1.7)	2 (2)	0 (0)	
Highest educational level, n (%)				
Elementary	22 (18.8)	19 (20)	3 (14)	.03[b]
High-school	45 (38.5)	41 (43)	4 (19)	
University <3 years	24 (20.5)	18 (19)	6 (29)	
University >3 years	26 (22.2)	18 (19)	8 (38)	
Marital status, n (%)				
Single	18 (15.4)	15 (16)	3 (14)	>.99
In relationship	99 (84.6)	81 (84)	18 (86)	
Country of birth, n (%)				
Sweden	96 (82.1)	81 (84)	15 (71)	.21
Other	21 (17.9)	15 (16)	6 (29)	
Children in the household, n (%)				
Yes	43 (36.8)	33 (34)	10 (48)	.32
No	74 (63.2)	63 (66)	11 (52)	
Current counseling, n (%)				
Yes	30 (25.6)	24 (25)	6 (29)	.78
No	87 (74.4)	72 (75)	15 (71)	
Clinical characteristics, mean (SD)				
HADS-A[c]	10.27 (2.94)	10.39 (3.11)	9.76 (2.00)	.25
HADS-D[d]	7.97 (3.15)	8.20 (3.26)	6.95 (2.42)	.05

[a]Employed versus other.

[b]Studied at university versus didn't study at university.

[c]HADS-A: Hospital Anxiety and Depression Scale-Anxiety.

[d]HADS-D: Hospital Anxiety and Depression Scale-Depression.

Results

Treatment Activity in the Randomized Controlled Trial

Of all participants allocated to intervention, 113 (96.6%, 113/117) initiated the introduction module, which was completed by 63 (53.9%, 63/117). Each of the remaining modules was completed by 7 or less of the participants. *Managing worry* and *Applied relaxation training* were the most frequently initiated and completed modules. The modules for *Communication training* and *Values in life* were not completed by any participant (Table 2).

Table 2. Number of participants in the randomized controlled trial (n=117) who initiated and completed the respective modules in the treatment program.

Module	Initiated, n (%)	Completed, n (%)
Introduction	113 (96.6)	63 (53.9)
Managing worry	23 (19.7)	7 (6.0)
Applied relaxation training	28 (24.0)	5 (4.3)
Behavioral activation	16 (13.7)	4 (3.4)
Fear and avoidance post myocardial infarction	7 (6.0)	3 (2.6)
Cognitive restructuring	11 (9.4)	2 (1.7)
Coping with insomnia	6 (5.1)	2(1.7)
Problem solving	4 (3.4)	2 (1.7)
Relapse prevention depression and anxiety	3 (2.6)	1 (0.9)
Communication skills	7 (6.0)	0 (0.0)
Values in life	3 (2.6)	0 (0.0)

Table 3. Total number of completed modules, completed assignments, and messages sent to therapist among the participants allocated to the intervention in the randomized controlled trial (n=117).

Number of completed modules, assignments, and sent internal messages at end of treatment period	Number of participants who completed modules, n (%)	Number of participants who completed assignments, n (%)	Number of participants who sent messages to therapist, n (%)
0	54 (46.2)	30 (25.6)	66 (56.4)
1	45 (38.5)	21 (17.9)	21 (17.9)
2	14 (12.0)	23 (20.5)	7 (6.0)
3	1 (0.9)	8 (6.8)	6 (5.1)
4	2 (1.7)	2 (3.4)	4 (3.4)
5	1 (0.9)	14 (12.0)	3 (2.6)
> 5	0 (0.0)	19 (16.2)	10 (8.5)

A minority of participants completed additional modules beyond the introductory module (18/117, 15.4%), completed more than 5 assignments (19/117, 16.2%), and sent more than 5 messages to the therapist (10/117, 8.5%; see Table 3).

The mean number of completed modules, completed assignments, and messages sent to therapist did not reach above 0.6 at any of the 14 treatment weeks. The total summed range for all 14 weeks was 0 to 5 for completed modules, 0 to 29 completed assignments, and 0 to 16 messages sent to therapist (Multimedia Appendix 6). Most assignments were completed during the first week of treatment (Figure 3). Over the course of treatment, the total number of completed assignments and messages sent to therapist declined. A slight increase in treatment activity was observed in the middle of the treatment period, which coincided with the collection of outcome assessments. A slight increase in completed assignments and messages sent to therapist was observed toward the end of the treatment period. A total number of 41 out of 117 (35.1%) participants opened one or more supplementary material or a video clip in the library. Among these participants, the average number of opened items was 3.85 (SD 4.55).

Telephone Interviews

We identified 4 main categories: (1) the portal, (2) the treatment program, (3) the therapist communication, and (4) the personal situation and required skills (Table 4). See Multimedia Appendix 7 for an expanded presentation of the qualitative results with the total number of participants who described experiences related to findings in respective category and illustrative quotes. In total, participants described 19 suggestions for improvement (Textbox 1).

Figure 3. Total number of completed modules, assignments, and messages sent to therapist during the 14-week treatment period.

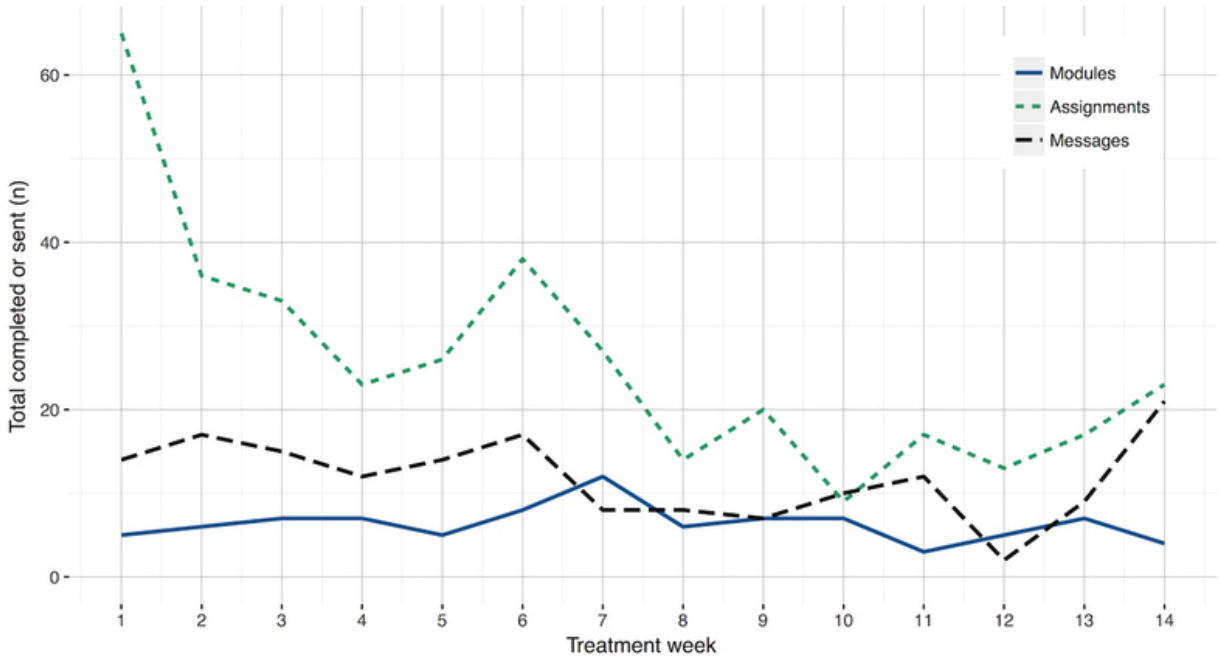


Table 4. Summary of positive and negative experiences described in interviews.

Category and subcategory	Findings	
	Positive experiences	Negative experiences
Portal		
Design	• Appealing interface with easy navigation	• Navigational difficulties, unfamiliar interface
Usability	• Easy and secure log-in procedure	• Complicated log-in procedure with technical failures • Required desktop or laptop, issues when using mobile device • Cumbersome to open PDF files
Treatment program		
Content of treatment material	• Relevant, well-written, and useful information	• Irrelevant outdated material and posts in discussion board • Repetitive material with poor readability
Working with the material	• Manageable difficulty, approach gave time to reflect • Time flexibility, possibility to select modules	• Strenuous, tedious, difficult, and time-consuming work • Too intensive work, restrictions in active modules felt rigid
Treatment period	• Deadline promoted activity toward end of treatment	• Treatment duration and time to work with modules was too short
Therapist communication		
Therapist feedback	• Tailored, available, and rapid feedback • Telephone conversations with therapist • Reminders were useful prompt to log in	• Lack of and irrelevant therapist feedback • Aversive and stressful reminders
Internet-based communication		• Preference for verbal and synchronous communication • Communication felt impersonal and involved a risk of misunderstanding
Personal situation and required skills		
Unpleasant emotions evoked by the intervention		• Bad conscience and guilt for being inactive • Treatment rekindled difficult memories and emotions • Fear of making mistakes
Lack of time		• Lack of time because of everyday life • Poor timing of treatment
Responding to outcome measures in questionnaires		• Questionnaires were difficult to understand, felt repetitive, and irrelevant • Strenuous work with questionnaires
Technical aspects		• Insufficient computer literacy • Intervention required time in front of a computer • Lack of Internet connection

Textbox 1. Summary of suggestions for improvement by category.

Portal

- Remove the completed modules to facilitate navigation
- Include the possibility to have several windows open at the same time
- Make the portal available via CD-ROM and as an app for mobile devices

Treatment program

- Have less focus on depression and anxiety following a myocardial infarction
- Include the possibility to ask medical questions to health professionals and other participants in the portal
- Include information concerning how to communicate with children
- Use easy-to-read language
- Use closed-ended questions with predetermined alternatives in the treatment program
- Prolong the treatment period and allow longer time for work with modules that feel relevant for the patient
- Make the program feel more fun for the intended users

Therapist communication

- Offer synchronous verbal therapist communication, via telephone calls
- Offer more therapist feedback in decision making concerning which modules to work with
- Use audio or video recordings of therapist feedback
- Entitle the patient with their name instead of username
- Include picture of the therapist in all conversations

Personal situation and required skills

- Individualize the outcome questionnaires
- Make the outcome questionnaires easier to understand
- Allow participants to access previous responses in the outcome questionnaires
- Offer access to treatment closer in time to the infarction

Discussion

Principal Findings

Our study focused on treatment activity and user experiences of an iCBT intervention to reduce symptoms of depression and anxiety among adults with a recent myocardial infarction. The results show that treatment activity was low with regard to completed modules and assignments and submitted internal messages to therapists. Various positive experiences, negative experiences, and suggestions for improvements were described in follow-up interviews related to the internet-based portal, treatment program, therapist communication, as well as the personal situation and required skills of the participants. Previous research shows inconclusive and variable results concerning treatment activity and user satisfaction of iCBT. Although some studies report high levels of adherence and sufficient treatment satisfaction [8], others indicate that adherence varies considerably between studies [27]. Furthermore, it has been suggested that internet interventions may be more attractive among individuals who are familiar with computers, express confidence in writing about thoughts and feelings, who are attracted to the opportunity to reflect, and who appreciate the anonymity provided by the medium [15]. Our findings strengthen these assumptions and indicate a need for more research to investigate for whom, when, and how iCBT interventions may be a suitable treatment alternative to effectively alleviate symptoms of depression and anxiety after a myocardial infarction.

Although iCBT shows promise as a mode of treatment for symptoms of depression and anxiety [8], few studies have investigated these interventions through clinical consecutive recruitment [38]. This study included patients recruited in routine cardiac care, which provides new insights regarding treatment activity and user experiences of internet-based interventions. Our findings illustrate that although most participants initiated treatment, few persisted with the iCBT treatment. This finding indicates that the intervention was unable to successfully motivate the users to engage in the treatment. Compared with consecutive recruitment, self-referral recruitment strategies have the potential to identify individuals who persist with iCBT and who find the treatment effective [39]. Thus, it has been suggested that iCBT may only be acceptable among a subgroup of patients [20]. Another possible explanation of the observed low treatment activity may be that the participants did not feel a need for psychological treatment. For example, symptoms of depression and anxiety may be perceived as a

normal reaction after a myocardial infarction. Patient attrition is an articulated issue for eHealth trials, which needs to be considered carefully when designing such interventions [26]. Our findings indicate a need for more research about how these patients experience a need for iCBT treatment, or if these types of interventions are better suited for certain subgroups of patients.

The observed low treatment activity and described negative experiences related to design and usability call attention to what has been described as a risk of distress and frustration when faced with technological difficulties [40]. In line with our findings, previous studies of iCBT interventions report that patients may experience struggles related to technology, delivery of treatment program, lack of support, and limited personalization of program content [41]. It is possible that our findings, in part, could be explained by the relatively high mean age among the participants. Higher age, as seen among patients with myocardial infarction [17], is associated with less use of the internet [42], low eHealth literacy [43], and unsuccessful skills needed to obtain reliable answers to health-related queries [44]. Moreover, older adults are more likely to report technological challenges in iCBT trials, and few studies have investigated user experiences of iCBT interventions for such populations [18]. The findings of this study illustrate the importance of efforts that aim to increase satisfaction and experienced usability among end users when developing eHealth interventions for patients with higher ages and a recent myocardial infarction.

Tailored interventions have the potential to successfully meet patient preferences by providing them with the choice of which treatment modules to work with [45] and adapting the treatment to the capacity of the patient [38]. In this study, the perceived positive and negative aspects varied considerably between individuals. This finding indicates a need to tailor interventions according to the intended end user's individual preferences, personal situation, as well as computer skills. For example, participants who perceive text-based material strenuous and time-consuming to read may benefit from a less-extensive version of the intervention. Participants who find it difficult to write about thought and feelings may benefit from the use of closed-ended questions with predetermined alternatives in the treatment program. Telephone calls may be offered as an alternative to written feedback to participants with a preference for verbal and synchronous communication. One potential way to tailor the content according to individual needs of the intended users is to use patient and public engagement during the development phase [46,47]. In this study, patients with experience of emotional distress after a myocardial infarction and cardiac nurses were consulted about their views on the treatment material in the later stages of development. Consultations are considered to be lower levels of patient and public participation, as it may quickly lead to insights but lacks a commitment to subsequent actions [33]. It is possible that a different approach, involving collaborations with patients during the whole development process, could have led to an intervention closer in line with the preferences of the intended end users. In light of our findings, we acknowledge the potential importance of using high degrees of patient and public involvement when developing iCBT interventions.

Limitations

In this study, there are methodological limitations that should be taken into consideration. The sample may not fully represent the population of patients with symptoms of depression and anxiety after a recent myocardial infarction. Patients were recruited in routine care at 25 Swedish cardiac clinics. Only patients below 75 years of age were invited to participate in the randomized trial. This may limit the generalizability and transferability with regard to older patients. The majority of the participants in the trial were males, employed, living in a relationship, and born in Sweden. Furthermore, only a subsample of those who took part in the intervention was interviewed. The reason for this was mainly practical, as we lacked necessary resources to collect qualitative data in the early stages of the study. This may imply a source of selection bias that may impact the results. We acknowledge that the qualitative results only reflect the experiences of a proportion of the whole sample in the RCT. Although the sample characteristics for participants in the RCT were represented in the sample of participants included in the follow-up telephone interviews, none of the participants who were unemployed or on sick leave were included in the follow-up interviews. Furthermore, a higher proportion of the interviewed participants had studied at a university compared with those who were not interviewed. This may imply a limited transferability to participants with lower levels of education. For example, it is possible that participants with experience of university studies may be more comfortable with text-based material and communication. Moreover, we did not collect any quantitative measure of computer literacy. Thus, we cannot make any claims about the actual computer literacy among the participants in our sample.

The data collection and analysis of the qualitative material may not fully represent the experiences of the interviewees. One psychologist who was not involved as a therapist in the treatment program conducted telephones interviews. Telephone interviews reduce the risk for socially desirable answers, may lead to increased sense of anonymity, and have the potential to make participants feel more comfortable [48,49]. On the other hand, telephone interviews make it impossible to observe nonverbal communication and create a comfortable physical setting where interviews take place [50]. We argue that the use of telephone communication and lack of previous contact with the interviewer promoted the participants to feel comfortable enough to be honest in their descriptions of their experiences and preferences. A semistructured interview guide with open-ended questions was used to cover our research questions, while still allowing for flexibility. The use of an interview guide implies instrumental consistency throughout the interviews [51,52]. Content analysis offers a systematic approach to describe patterns in text-based data [53,54]. However, there is always an embedded element of abstraction in qualitative analyses, which is impossible to completely disregard [51]. Thus, 2 authors with different backgrounds analyzed the data. We acknowledge that it is possible that potentially valuable information may have been lost due to potential biases or preconceptions.

Suggestions for Future Research

The findings indicate a need for rigorous preparations before conducting iCBT interventions for adults with depression or anxiety after a recent myocardial infarction. There is a need for future research that investigates ways to ensure that development of these interventions is more adapted to the intended end users. The low treatment activity and negative experiences related to the use of the internet platform and the treatment content call attention to the importance of usability and feasibility trials. Future research should investigate patient, therapist, and treatment-related factors to improve treatment activity in internet-based interventions implemented in this population.

Conclusions

Patients with symptoms of depression and anxiety after a recent myocardial infarction showed low treatment activity in guided iCBT with regard to completed modules, assignments, and messages sent to their therapist. They describe various negative experiences and suggestions for improvement, calling attention to the need for researchers to carefully consider the preferences, personal situation, and required skills of the end users during the development of these interventions. The findings indicate several challenges that need to be addressed to improve treatment activity, user satisfaction, and usability of internet interventions in this population.

Acknowledgments

This study is part of U-CARE strategic research environment funded by the Swedish Research Council (dnr 2009–1093). U-CARE Heart is also funded by the Swedish Heart and Lung Association, the Uppsala-Örebro Regional Research Council, Swedish Research Council for Health, Working Life, and Welfare (dnr 2014–4947), and the Vårdal foundation (dnr 2014–0114). The funders had no involvement in the study conception, data collection, analysis, and manuscript production.

Authors' Contributions

EW, FN, and GB conceived and designed the study. GB collected the data. EM and TC analyzed the data and drafted the manuscript. FN, EO, CH, and GB read and revised the draft. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Screenshot of the U-CARE Heart portal.

[JPG File, 2MB - jmir_v20i3e87_app1.jpg]

Multimedia Appendix 2

Description of development process.

[PDF File (Adobe PDF File), 14KB - jmir_v20i3e87_app2.pdf]

Multimedia Appendix 3

Interview guide.

[PDF File (Adobe PDF File), 18KB - jmir_v20i3e87_app3.pdf]

Multimedia Appendix 4

Examples of the steps in the qualitative analysis.

[PDF File (Adobe PDF File), 15KB - jmir_v20i3e87_app4.pdf]

Multimedia Appendix 5

Backgrounds of the researchers involved in data collection and qualitative analysis.

[PDF File (Adobe PDF File), 8KB - jmir_v20i3e87_app5.pdf]

Multimedia Appendix 6

Completed modules, completed assignments, and messages sent to therapist for each treatment week of the intervention.

[PDF File (Adobe PDF File), 14KB - jmir_v20i3e87_app6.pdf]

Multimedia Appendix 7

An expanded presentation of the qualitative findings with illustrative quotes from interviews.

[PDF File (Adobe PDF File), 54KB - jmir_v20i3e87_app7.pdf]

References

1. Lane D, Carroll D, Ring C, Beevers DG, Lip GY. The prevalence and persistence of depression and anxiety following myocardial infarction. Br J Health Psychol 2002 Feb;7(Pt 1):11-21. [doi: 10.1348/135910702169321] [Medline: 14596714]
2. Thombs BD, Bass EB, Ford DE, Stewart KJ, Tsilidis KK, Patel U, et al. Prevalence of depression in survivors of acute myocardial infarction. J Gen Intern Med 2006 Jan;21(1):30-38 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.00269.x] [Medline: 16423120]
3. Meijer A, Conradi HJ, Bos EH, Thombs BD, van Melle JP, de Jonge P. Prognostic association of depression following myocardial infarction with mortality and cardiovascular events: a meta-analysis of 25 years of research. Gen Hosp Psychiatry 2011;33(3):203-216 [FREE Full text] [doi: 10.1016/j.genhosppsych.2011.02.007] [Medline: 21601716]
4. Roest AM, Martens EJ, Denollet J, de Jonge P. Prognostic association of anxiety post myocardial infarction with mortality and new cardiac events: a meta-analysis. Psychosom Med 2010 Jul;72(6):563-569. [doi: 10.1097/PSY.0b013e3181dbff97] [Medline: 20410247]
5. DiMatteo MR, Lepper HS, Croghan TW. Depression is a risk factor for noncompliance with medical treatment: meta-analysis of the effects of anxiety and depression on patient adherence. Arch Intern Med 2000 Jul 24;160(14):2101-2107. [Medline: 10904452]
6. Bäck M, Cider Å, Herlitz J, Lundberg M, Jansson B. The impact on kinesiophobia (fear of movement) by clinical variables for patients with coronary artery disease. Int J Cardiol 2013 Jul 31;167(2):391-397. [doi: 10.1016/j.ijcard.2011.12.107] [Medline: 22305808]
7. Benyamini Y, Roziner I, Goldbourt U, Drory Y, Gerber Y, Israel Study Group on First Acute Myocardial Infarction. Depression and anxiety following myocardial infarction and their inverse associations with future health behaviors and quality of life. Ann Behav Med 2013 Dec;46(3):310-321. [doi: 10.1007/s12160-013-9509-3] [Medline: 23645421]
8. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
9. Arnberg FK, Linton SJ, Hultcrantz M, Heintz E, Jonsson U. Internet-delivered psychological treatments for mood and anxiety disorders: a systematic review of their efficacy, safety, and cost-effectiveness. PLoS One 2014;9(5):e98118 [FREE Full text] [doi: 10.1371/journal.pone.0098118] [Medline: 24844847]
10. van Beugen S, Ferwerda M, Hoeve D, Rovers MM, Spillekom-van Koulil S, van Middendorp H, et al. Internet-based cognitive behavioral therapy for patients with chronic somatic conditions: a meta-analytic review. J Med Internet Res 2014;16(3):e88 [FREE Full text] [doi: 10.2196/jmir.2777] [Medline: 24675372]
11. Cowie MR, Chronaki CE, Vardas P. e-Health innovation: time for engagement with the cardiology community. Eur Heart J 2013 Jul;34(25):1864-1868. [doi: 10.1093/eurheartj/ehs153] [Medline: 22733834]
12. Glozier N, Christensen H, Naismith S, Cockayne N, Donkin L, Neal B, et al. Internet-delivered cognitive behavioural therapy for adults with mild to moderate depression and high cardiovascular disease risks: a randomised attention-controlled trial. PLoS One 2013;8(3):e59139 [FREE Full text] [doi: 10.1371/journal.pone.0059139] [Medline: 23555624]
13. Andersson G. Internet-delivered psychological treatments. Annu Rev Clin Psychol 2016;12:157-179. [doi: 10.1146/annurev-clinpsy-021815-093006] [Medline: 26652054]
14. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the internet? A systematic review of the published literature. J Med Internet Res 2006;8(2):e10 [FREE Full text] [doi: 10.2196/jmir.8.2.e10] [Medline: 16867965]
15. Beattie A, Shaw A, Kaur S, Kessler D. Primary-care patients' expectations and experiences of online cognitive behavioural therapy for depression: a qualitative study. Health Expect 2009 Mar;12(1):45-59. [doi: 10.1111/j.1369-7625.2008.00531.x] [Medline: 19250152]
16. 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]
17. Anand SS, Islam S, Rosengren A, Franzosi MG, Steyn K, Yusufali AH, INTERHEART Investigators. Risk factors for myocardial infarction in women and men: insights from the INTERHEART study. Eur Heart J 2008 Apr;29(7):932-940. [doi: 10.1093/eurheartj/ehn018] [Medline: 18334475]
18. Crabb RM, Cavanagh K, Proudfoot J, Learmonth D, Rafie S, Weingardt KR. Is computerized cognitive-behavioural therapy a treatment option for depression in late-life? A systematic review. Br J Clin Psychol 2012 Nov;51(4):459-464. [doi: 10.1111/j.2044-8260.2012.02038.x] [Medline: 23078214]
19. Klein JP, Gamon C, Späth C, Berger T, Meyer B, Hohagen F, et al. Does recruitment source moderate treatment effectiveness? A subgroup analysis from the EVIDENT study, a randomised controlled trial of an internet intervention for depressive

symptoms. Br Med J Open 2017 Jul 13;7(7):e015391 [FREE Full text] [doi: 10.1136/bmjopen-2016-015391] [Medline: 28710212]
20. 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]
21. Kazdin A. Perceived barriers to treatment participation and treatment acceptability among antisocial children and their families. J Child Fam Stud 2000;9(2):157-174. [doi: 10.1023/A:100941490]
22. Gun SY, Titov N, Andrews G. Acceptability of Internet treatment of anxiety and depression. Australas Psychiatry 2011 Jun;19(3):259-264. [doi: 10.3109/10398562.2011.562295] [Medline: 21682626]
23. Musiat P, Goldstone P, Tarrier N. Understanding the acceptability of e-mental health--attitudes and expectations towards computerised self-help treatments for mental health problems. BMC Psychiatry 2014;14:109 [FREE Full text] [doi: 10.1186/1471-244X-14-109] [Medline: 24725765]
24. Wallin EE, Mattsson S, Olsson EM. The preference for internet-based psychological interventions by individuals without past or current use of mental health treatment delivered online: a survey study with mixed-methods analysis. JMIR Ment Health 2016;3(2):e25 [FREE Full text] [doi: 10.2196/mental.5324] [Medline: 27302200]
25. Kaltenthaler E, Sutcliffe P, Parry G, Beverley C, Rees A, Ferriter M. The acceptability to patients of computerized cognitive behaviour therapy for depression: a systematic review. Psychol Med 2008 Nov;38(11):1521-1530. [doi: 10.1017/S0033291707002607] [Medline: 18205964]
26. 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]
27. Păsărelu CR, Andersson G, Bergman NL, Dobrean A. Internet-delivered transdiagnostic and tailored cognitive behavioral therapy for anxiety and depression: a systematic review and meta-analysis of randomized controlled trials. Cogn Behav Ther 2017 Jan;46(1):1-28. [doi: 10.1080/16506073.2016.1231219] [Medline: 27712544]
28. 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]
29. Amichai-Hamburger Y, Klomek AB, Friedman D, Zuckerman O, Shani-Sherman T. The future of online therapy. Comput Human Behav 2014 Dec;41:288-294. [doi: 10.1016/j.chb.2014.09.016]
30. McMain S, Newman MG, Segal ZV, DeRubeis RJ. Cognitive behavioral therapy: current status and future research directions. Psychother Res 2015;25(3):321-329. [doi: 10.1080/10503307.2014.1002440] [Medline: 25689506]
31. Norlund F, Wallin E, Olsson E, Wallert J, Burell G, von Essen L, et al. Internet-based cognitive behavior therapy for symptoms of depression and anxiety among patients with a recent myocardial infarction: The U-CARE Heart randomized trial. J Med Internet Res 2018;(forthcoming) (forthcoming). [doi: 10.2196/jmir.9710]
32. Ericsson KA, Simon HA. How to study thinking in everyday life: contrasting think-aloud protocols with descriptions and explanations of thinking. Mind Cult Activ 1998 Jul;5(3):178-186. [doi: 10.1207/s15327884mca0503_3]
33. Oliver SR, Rees RW, Clarke-Jones L, Milne R, Oakley AR, Gabbay J, et al. A multidimensional conceptual framework for analysing public involvement in health services research. Health Expect 2008 Mar;11(1):72-84. [doi: 10.1111/j.1369-7625.2007.00476.x] [Medline: 18275404]
34. Norlund F, Olsson EM, Burell G, Wallin E, Held C. Treatment of depression and anxiety with internet-based cognitive behavior therapy in patients with a recent myocardial infarction (U-CARE Heart): study protocol for a randomized controlled trial. Trials 2015 Apr 11;16:154 [FREE Full text] [doi: 10.1186/s13063-015-0689-y] [Medline: 25873137]
35. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [Medline: 6880820]
36. Svanborg P, Asberg M. A comparison between the Beck Depression Inventory (BDI) and the self-rating version of the Montgomery Asberg Depression Rating Scale (MADRS). J Affect Disord 2001 May;64(2-3):203-216. [Medline: 11313087]
37. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
38. 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]
39. Cuijpers P, van Straten A, Andersson G. Internet-administered cognitive behavior therapy for health problems: a systematic review. J Behav Med 2008 Apr;31(2):169-177 [FREE Full text] [doi: 10.1007/s10865-007-9144-1] [Medline: 18165893]
40. Rozental A, Boettcher J, Andersson G, Schmidt B, Carlbring P. Negative effects of internet interventions: a qualitative content analysis of patients' experiences with treatments delivered online. Cogn Behav Ther 2015;44(3):223-236. [doi: 10.1080/16506073.2015.1008033] [Medline: 25705924]
41. Knowles SE, Lovell K, Bower P, Gilbody S, Littlewood E, Lester H. Patient experience of computerised therapy for depression in primary care. Br Med J Open 2015 Nov 30;5(11):e008581. [doi: 10.1136/bmjopen-2015-008581] [Medline: 26621513]

42. Kummervold PE, Chronaki CE, Lausen B, Prokosch H, Rasmussen J, Santana S, et al. eHealth trends in Europe 2005-2007: a population-based survey. J Med Internet Res 2008;10(4):e42 [FREE Full text] [doi: 10.2196/jmir.1023] [Medline: 19017584]
43. Richtering SS, Hyun K, Neubeck L, Coorey G, Chalmers J, Usherwood T, et al. eHealth literacy: predictors in a population with moderate-to-high cardiovascular risk. JMIR Hum Factors 2017 Jan 27;4(1):e4 [FREE Full text] [doi: 10.2196/humanfactors.6217] [Medline: 28130203]
44. Agree EM, King AC, Castro CM, Wiley A, Borzekowski DLG. "It's got to be on this page": age and cognitive style in a study of online health information seeking. J Med Internet Res 2015;17(3):e79 [FREE Full text] [doi: 10.2196/jmir.3352] [Medline: 25831483]
45. Andersson G, Estling F, Jakobsson E, Cuijpers P, Carlbring P. Can the patient decide which modules to endorse? An open trial of tailored internet treatment of anxiety disorders. Cogn Behav Ther 2011;40(1):57-64. [doi: 10.1080/16506073.2010.529457] [Medline: 21337215]
46. Brett J, Staniszewska S, Mockford C, Herron-Marx S, Hughes J, Tysall C, et al. A systematic review of the impact of patient and public involvement on service users, researchers and communities. Patient 2014;7(4):387-395. [doi: 10.1007/s40271-014-0065-0] [Medline: 25034612]
47. Entwistle VA, Renfrew MJ, Yearley S, Forrester J, Lamont T. Lay perspectives: advantages for health research. Br Med J 1998 Feb 07;316(7129):463-466 [FREE Full text] [Medline: 9492683]
48. Opdenakker RJ. 2006. Advantages and disadvantages of four interview techniques in qualitative research URL: https://pure.tue.nl/ws/files/1948695/Metis202565.pdf [accessed 2018-02-27] [WebCite Cache ID 6xYDFioIo]
49. Novick G. Is there a bias against telephone interviews in qualitative research? Res Nurs Health 2008 Aug;31(4):391-398 [FREE Full text] [doi: 10.1002/nur.20259] [Medline: 18203128]
50. Sturges JE, Hanrahan KJ. Comparing telephone and face-to-face qualitative interviewing: a research note. Qual Res 2016 Aug 15;4(1):107-118. [doi: 10.1177/1468794104041110]
51. Patton MQ. Qualitative Research and Evaluation Methods. 3rd edition. London: SAGE; 2002.
52. Lincoln YS, Guba EG. Naturalistic Inquiry. London: SAGE Publications, Inc; 1985.
53. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
54. Sandelowski M. Whatever happened to qualitative description? Res Nurs Health 2000 Aug;23(4):334-340. [Medline: 10940958]

Abbreviations

HADS: Hospital Anxiety and Depression Scale
HADS-A: Hospital Anxiety and Depression Scale-Anxiety
HADS-D: Hospital Anxiety and Depression Scale-Depression.
iCBT: Internet-based cognitive behavioral therapy
MADRS-S: Montgomery Asberg Depression Rating Scale Short form
RCT: randomized controlled trial
SMS: short message service

Edited by G Eysenbach; submitted 19.12.17; peer-reviewed by T Poder, P Musiat, A Parks; comments to author 18.01.18; revised version received 05.02.18; accepted 08.02.18; published 16.03.18.

Please cite as:
Wallin E, Norlund F, Olsson EMG, Burell G, Held C, Carlsson T
Treatment Activity, User Satisfaction, and Experienced Usability of Internet-Based Cognitive Behavioral Therapy for Adults With Depression and Anxiety After a Myocardial Infarction: Mixed-Methods Study
J Med Internet Res 2018;20(3):e87
URL: http://www.jmir.org/2018/3/e87/
doi:10.2196/jmir.9690
PMID:29549067

Original Paper

A Mobile App to Improve Self-Management of Individuals With Type 2 Diabetes: Qualitative Realist Evaluation

Laura Desveaux[1,2], PT, PhD; James Shaw[1,2], PT, PhD; Marianne Saragosa[1], RN, MN; Charlene Soobiah[1,2], HBSc; Husayn Marani[1], MSc; Jennifer Hensel[1], MSc, MD; Payal Agarwal[1], BASc, MD; Nike Onabajo[1], PMP, MSc; R Sacha Bhatia[1,2], MBA, MD; Lianne Jeffs[3,4], RN, MSc, PhD

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

[2]Institute for Health Policy, Management, and Evaluation, University of Toronto, Toronto, ON, Canada

[3]Keenan Research Centre, Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, ON, Canada

[4]American Academy of Nursing, Washington, DC, United States

Corresponding Author:
Laura Desveaux, PT, PhD
Institute for Health System Solutions and Virtual Care
Women's College Hospital
76 Grenville St
Toronto, ON, M5S 1B2
Canada
Phone: 1 416 323 6400 ext 8356
Email: laura.desveaux@wchospital.ca

Abstract

Background: The increasing use of Web-based solutions for health prevention and promotion presents opportunities to improve self-management and adherence to guideline-based therapy for individuals with type 2 diabetes (T2DM). Despite promising preliminary evidence, many users stop using Web-based solutions due to the burden of data entry, hidden costs, loss of interest, and a lack of comprehensive features. Evaluations tend to focus on effectiveness or impact and fail to evaluate the nuanced variables that may interact to contribute to outcome success (or failure).

Objective: This study aimed to evaluate a Web-based solution for improving self-management in T2DM to identify key combinations of contextual variables and mechanisms of action that explain for whom the solution worked best and in what circumstances.

Methods: A qualitative realist evaluation was conducted with one-on-one, semistructured telephonic interviews completed at baseline, and again toward the end of the intervention period (3 months). Topics included participants' experiences of using the Web-based solution, barriers and facilitators of self-management, and barriers and facilitators to effective use. Transcripts were analyzed using thematic analysis strategies, after which the key themes were used to develop statements of the relationships between the key contextual factors, mechanisms of action, and impact on the primary outcome (glycated hemoglobin, HbA_{1c}).

Results: Twenty-six interviews (14 baseline, 12 follow-up) were completed with 16 participants with T2DM, and the following 3 key groups emerged: the easiest fit, the best fit, and those who failed to activate. Self-efficacy and willingness to engage with the solution facilitated improvement in HbA_{1c}, whereas competing priorities and psychosocial issues created barriers to engagement. Individuals with high baseline self-efficacy who were motivated, took ownership for their actions, and prioritized diabetes management were early and eager adopters of the app and recorded improvements in HbA_{1c} over the intervention period. Individuals with moderate baseline self-efficacy and no competing priorities, who identified gaps in understanding of how their actions influence their health, were slow to adopt use but recorded the greatest improvements in HbA_{1c}. The final group had low baseline self-efficacy and identified a range of psychosocial issues and competing priorities. These participants were uncertain of the benefits of using a Web-based solution to support self-management, ultimately resulting in minimal engagement and no improvement in HbA_{1c}.

Conclusions: Self-efficacy, competing priorities, previous behavior change, and beliefs about Web-based solutions interact to determine engagement and impact on the clinical outcomes. Considering the balance of these patient characteristics is likely to help health care providers identify individuals who are apt to benefit from a Web-based solution to support self-management of

T2DM. Web-based solutions could be modified to incorporate the existing screening measures to identify individuals who are at risk of suboptimal adherence to inform the provision of additional support(s) as needed.

(J Med Internet Res 2018;20(3):e81) doi:10.2196/jmir.8712

KEYWORDS

telemedicine; diabetes mellitus; self-management; qualitative research

Introduction

The number of people affected with diabetes worldwide has increased from 171 million to 422 million between 2000 and 2017, making it one of the most costly and devastating chronic diseases [1]. In Ontario, Canada, the prevalence of diabetes increased by 69% from 1995 to 2005 [2], exceeding the global increase of 60% previously projected to occur between 1995 and 2030 [3,4]. This dramatic rise is attributed to new cases of type 2 diabetes mellitus (T2DM) [5,6], driven by increasing rates of obesity [7,8]. Physical inactivity, smoking, alcohol consumption, and poor dietary habits have also been identified as risk factors that significantly increase an individual's risk of developing T2DM [9]. Long-term complications include retinopathy, nephropathy, autonomic neuropathy leading to cardiovascular symptoms, and peripheral neuropathy with a risk of foot ulcers and amputations [5]. Most notably, individuals who have T2DM are twice as likely to die over a 12 year follow-up period compared with those without diabetes [10,11].

Given the severity and nature of disease progression, a cornerstone of clinical management is the process of teaching individuals how to manage their diabetes. An individual's attitudes, beliefs, and knowledge about diabetes may affect diabetes self-management, including their adherence to prescribed pharmacotherapy [12,13], highlighting the need for individualized, patient-centered approaches. T2DM education and self-management education is a cost-effective approach [14] and has a direct impact on patients' glycemic control [15]. Duration of contact between educator and patient has been noted to be a significant predictor of improved glycemic control in this population, underscoring the role of social support [15].

Despite advances in diabetes treatment and education, adherence to diabetes self-management regimens continues to be the most significant determinant of achieving clinical targets (ie, glycemic control) [16,17]. Barriers to diabetes management include individual attitudes and beliefs, knowledge, financial constraints, and social support [18-21]. Given the prevalence of mobile devices [22] and the increasing use of Web-based solutions for health prevention and promotion [23,24], mobile phone apps have emerged as a potential solution to improve self-management and adherence to guideline-based therapy due to their accessibility, low cost, and interactive potential [22]. These apps may include a range of features, including blood glucose monitoring, medication tracking, exercise tracking, and dietary management [25,26]. Although preliminary evidence looks promising [27-29], many app users stop using health apps due to high burden of data entry, hidden costs, loss of interest, and a lack of comprehensive features in a single solution [23,30-32]. Formal evaluations of Web-based solutions tend to focus on the effectiveness or impact and fail to evaluate the nuanced variables that may contribute to success (or failure) [33]. To address this gap in the literature, we conducted a qualitative realist evaluation as part of a larger randomized trial of a mobile-based self-management app to improve self-management in T2DM [34]. The objective was to identify key combinations of contextual variables and mechanisms of action that explain for whom the app worked best and in what circumstances.

Methods

Study Design

A qualitative realist evaluation [35] was embedded as part of a larger pragmatic, randomized, wait-list control trial to evaluate whether and how a mobile app designed to improve self-management and experience of care among patients with T2DM [34]. Realist evaluation is a methodology used to unpack the black box of implementation [36] by purposively examining the actions required by those involved in an intervention to ensure its success, including participants and those involved in implementation. This methodology enables a rigorous assessment of the contextual influences and strategies by which the intervention is adopted or rejected, enabling and understanding of how and why the implementation succeeds or fails. Specifically, a realist evaluation provides an explanation for why study outcomes occur, involves multimethods involving quantitative and qualitative approaches, and uses a theory-driven approach that guides the study design [35].

Trial participants were randomized to either an immediate treatment group (ITG) or a wait-list control group (WLC). The ITG group began using the mobile app immediately for a duration of 3 months. A series of quantitative outcomes were collected as part of the trial and are outlined in the original protocol [34]. Of particular relevance to this realist evaluation were the following 2 outcomes: glucose control (measured by HbA_{1c}), and the Problem Areas in Diabetes 5 (PAID 5) [37], a measure of disease-specific self-efficacy that emphasizes well-being. Self-efficacy refers to an individual's belief in his or her capability to achieve a given objective, which is a well-established mediator of health behaviors [38].

The intervention was implemented and supported by the Ontario Telemedicine Network (OTN), a nonprofit, government-funded organization and the largest provider of telemedicine services in the province of Ontario [39]. The protocol received ethics approval from Research Ethics Boards at participating institutions, including Women's College Hospital, St. Joseph's Care Group, North York General Hospital, and William Osler Health System. The larger trial is registered on ClinicalTrials.gov (NCT02813343).

Intervention

The intervention is a commercially available app designed to serve as a Web-based coach for patients with T2DM (WellDoc Bluestar allows participants to enter a range of baseline clinical information, in addition to ongoing data related to diabetes management, including blood glucose values, daily medications, food intake, and activity levels). The app analyzes inputted data to provide tailored messaging to coach participants with respect to their diabetes management. Participants also had the option of emailing a SMART Visit report to a member of their care team via the app, which provides them with an overview of inputted data over a period specified by the participant. At the time of this study, the app did not include secure messaging with providers or social functionality to connect participants with one another. The mobile app has been shown to improve glycemic control (as represented by reduction in levels of HbA_{1c}) in other contexts and settings [28,40].

The Web-based solution was implemented across 3 Diabetes Education Centers selected by the OTN. These sites were the Diabetes Health Centre in Thunder Bay, the Diabetes Education Center at North York General Hospital, and 2 Diabetes Education Centers belonging to the William Osler Health System. The OTN provided each site with funding for a site project coordinator who was responsible for recruiting participants and introducing them to the app. More than 4500 patients are seen across these sites annually, representing a socially and ethnically diverse group of individuals with diabetes. Each site serves distinct populations, including a large Indigenous population in Thunder Bay and visible minorities and newcomers in the William Osler Health System.

Recruitment

Potential interview participants were recruited from the ITG group to ensure maximum potential for exposure to the app. The complete recruitment strategy has been described previously [34]. One-on-one, semistructured telephonic interviews were conducted, with questions guided by the principles of Realist Evaluation [35]. Topics include participants' experiences of learning about and using the technology, barriers and facilitators of self-management, and barriers and facilitators to effective use (refer to Multimedia Appendix 1 for interview questions). Participants were interviewed at baseline and again toward the completion of the intervention period (3 months).

Data Analysis

Interviews were conducted by an experienced qualitative researcher, audio-recorded, and transcribed by a third party. Transcripts were analyzed using thematic analysis strategies [35,41], which included identifying key themes that demonstrate important contextual influences and mechanisms of action for the Web-based solution in real-world health care settings. Recruitment continued until data saturation was reached. A minimum of 2 reviewers independently coded all transcripts using an open coding process. Following the first 5 interviews, a coding schema was created to guide the analysis of the subsequent interviews. Open coding was applied throughout the analysis for content that did not fit within the existing coding schema. Consolidation of codes was achieved through consultation with a third reviewer. There were no disagreements with respect to coding.

Several strategies were employed to ensure credibility of the data, such as using multiple sources of data, having key collaborators participate in the triangulation analysis and the return of findings (construct and external validity), examining points of convergence and divergence within and across cases (internal validity through cross comparative analyses), and having a stepped analysis process whereby there is an initial independent review of the data by 3 reviewers (LD, MS, and LJ) who then met to reach consensus on the common themes (reliability) [42].

After the thematic analyses of all qualitative data had been completed, the key themes identified were used to develop statements of the relationships between (1) key contextual factors, (2) the mechanisms by which they affect the implementation of the Web-based solution, and (3) the impact on the outcomes of the intervention itself (in Realist Evaluation these statements are referred to as *Context-Mechanism-Outcome [C-M-O] Configurations*) [35].

Results

Findings

A total of 26 interviews (14 baseline, 12 follow-up) were completed with 16 participants with T2DM across the 3 sites. Of the 14 participants who completed the baseline interview, 3 had dropped out and 1 was unavailable at follow-up; therefore, additional 2 participants were recruited to achieve data saturation at follow-up. Characteristics of patient participants are shown in Table 1. Patient participants were grouped according to their primary outcome from the trial data (HbA_{1c}) and self-reported level of engagement with the app to describe C-M-O configurations. From the data, 3 groups emerged that are described in Table 1 (see Table 1 for a summary of participant characteristics by group).

Group 1: The Easiest Fit—Engaged Early Adopters

The first C-M-O configuration concerns a group of individuals who had a high level of self-efficacy when self-managing their T2DM before the intervention, took ownership for their actions and were motivated to change, reported no competing priorities, were keen to engage with mobile technology to help support self-management, and were newly diagnosed with T2DM (3-9 months). A high level of preintervention self-efficacy was illustrated by having a positive mindset and reported behavior change and HbA_{1c} levels that were improving before enrolment in the study. These actions included increased physical activity, portion control, avoiding high-fat foods, and medication adherence.

Table 1. Patient participant characteristics.

Characteristics	Group 1: Engaged, early adopters (n=4)	Group 2: Engaged, slow adopters (n=4)	Group 3	
			Low engagers (n=5)	Dropouts (n=3)
Age in years, mean (range)	57 (51-63)	59 (49-67)	42 (32-52)	45 (37-49)
Sex (male:female)	4:0	2:2	2:3	2:1
Time since T2DM[a] diagnosis, mean (range)	6 months (3-9 months)	19 years (9-26 years)	6 years (4 months-13 years)	3 years (3 months-27 years)
Complications from T2DM, n	1	3	2	4
Site, n				
NYGH[b]	0	3	1	1
Thunder Bay	2	1	1	2
WOHS[c]	2	0	3	0
Marital status, n				
Married	3	2	2	-
Divorced	1	2	-	1
Common law	-	-	1	1
Single	1	1	2	-
Ethnic background, n				
White	2	3	4	-
First Nations	-	-	-	2
African American	-	1	1	1
Highest education, n				
High school	-	3	-	-
University	1	-	2	2
Postgraduate	3	1	3	1
Type of therapy, n				
Oral	4	-	3	1
Insulin	-	1	1	-
Combination	-	3	1	2

[a]T2DM: type 2 diabetes.

[b]NYGH: North York General Hospital.

[c]WOHS: William Osler Health System.

Individuals in this group also exhibited proactive, information-seeking behavior to help them manage their T2DM:

> *I started looking up what information was on [the Diabetes Canada website], found out about what the daily types of meals you should have to control your blood sugar, and I immediately started following that.* [SITE A05]

Participants described various aspects of their identity that fuelled their motivation to change, which involved social activities, occupation, family, and overall quality of life. Narratives outlined feelings of accountability to oneself and taking ownership for individual actions:

> *Well it's to my own benefit right, you know, if I do it then I'm the one that hopefully gets rewarded, it's self-serving in a lot of places for me right...I, you know, I want to still have life left and do things I want to do.* [SITE A05]

Unlike participants in other groups, high engagers did not identify competing priorities that interfered with their ability to manage their T2DM. Instead, they described integrating management strategies into the existing schedules and routines. The majority described strong support networks at home that helped them adhere to their prescribed diet and activity recommendations. Only 1 individual described his social network and relationship with his family, but did not link these relationships to his self-management behavior. The study project coordinator, who was responsible for introducing the app to participants, was also viewed as a source of self-management support at 1 particular site:

> *We kinda have a kind of conversation when we get together...more like a friendly visit instead of an 'oh I have to report to the nurse'...Yeah, I think I like her, and I know that I can just phone her up, she told me that, just phone her up, I mean if you need any advice or whatever–so that was nice.* [SITE B06]

These individuals were interested in using mobile technology to improve their health and enthusiastically engaged with the app immediately and consistently thereafter. Engaging with the intervention led to the activation of several mechanisms of change for these individuals. The data entry requirement of the app reinforced and strengthened the pre-existing accountability to self. Individuals described feeling "grounded" and "honest," and explained how visualizations of their data helped to keep them on track and triggered a greater commitment to self-management:

> *[The application] keeps me grounded and keeps me honest. Even though you can put whatever you want in there and say 'Oh yeah I'm having 10 slices of pizza and you only put in 1'...But it keeps me honest...It keeps me on track.* [SITE B01]

Data entry led to positive performance feedback that further enhanced the individual's ability to self-monitor. Visual feedback displaying desired outcomes "reinforced positive behaviors" and encouraged participants to continue making informed choices and monitoring outcomes. These mechanisms interacted with each individual's context to produce improvements in their primary clinical outcome (HbA_{1c}) and their overall ability to manage their T2DM. These self-reported improvements were supported by the quantitative outcomes collected as part of the larger trial (refer to Table 2).

> *My A_{1c} is now pretty perfect...In April it was 10.4 and uhm, in November it was 6.5.* [SITE B06]

All individuals in this group highlighted that, in addition to the primary outcomes highlighted above, engaging with the app helped increase their awareness of their T2DM and the impact of stress and diet on their glucose readings. This increased awareness translated into increased self-efficacy with respect to self-managing T2DM.

> *Before I was clueless, not totally clueless, I just thought it was the sugary thing, I didn't know how much the carbs got involved, and the fats.* [SITE B 06]

Group 2: The Best Fit—Engaged Slow Adopters

The second group of C-M-O configurations concerns a group of individuals who had moderate self-efficacy in terms of self-managing their T2DM before the intervention, described an incomplete understanding of how actions influence their health and why, reported no competing priorities, were open to the idea of using mobile technology to support self-management, and had a long-standing diagnosis of T2DM.

Participants in this group described frustrations with the episodic nature of managing their condition and repeated unsuccessful attempts to "fine-tune" their self-management strategy. Despite a partial understanding of strategies to manage T2DM, these individuals strongly expressed their desire to fill these knowledge gaps. As a result, participants in this group identified the need for a specific solution that targeted their ability to achieve a more nuanced understanding of their T2DM (9-26 years):

> *You just go through stages of depression, you go through stages of anger, depression, denial, and then you sort of wake up and say, 'Ok, I'll just keep trying.' And then you try again and you're good for a couple years and then something happens, you get sick or, and there sometimes it's discouraging because...it doesn't matter what you do. I can take all the insulin I want but for me, when I'm sick, I can't get my blood sugar down.* [SITE A02]

> *I have a feeling that my readings after dinner are still too high, but because I can't break it down I don't have the motivation to take the last step which is to write down everything you eat at dinner for the next three weeks and how much insulin you took so then we can address the little problems.* [SITE C02]

Participants described how the mobile app met these needs, which varied depending on the individual participant. Overall, the intervention enabled participants to track inputs such as diet and stress and their impact on a specific outcome, blood glucose levels. This mechanism of performance feedback increased participant awareness around which actions influenced their disease management:

> *With this, looking at whether I'm putting in my carbs and that, thinking 'Ok, well I can only have...this. Yeah. That's all I can have.' So I think it's made me more think about the fact that I can only eat so much and before it was just like, ah the heck with it.* [SITE B02]

Data tracking and trend visualization increased participants' sense of accountability for their actions. Participants described their new-found accountability both to themselves and their health care providers, to whom they were very well connected. Data visualization enabled participants to see positive results in-between health care provider visits and encouraged incremental increases in engagement with the app over time:

> *The app helps me, you know, to be testing my blood and recording it and seeing any positive changes that I'm making. And the positive changes in turn help to sort of encourage me to continue it...so it's like a circle if you will.* [SITE C03]

Real-time, nuanced performance feedback displayed glucose readings alongside symptom and dietary inputs and organized inputs by time of day. This enabled participants' self-monitoring ability. In the case of this group, feedback displaying desired outcomes "drew attention to positive behaviors" and encouraged participants to continue making healthy decisions and monitoring outcomes. These mechanisms interacted with each individual's context to produce improvements in their primary clinical outcome (HbA_{1c}) and their overall ability to manage their T2DM.

Table 2. Mixed-methods results matrix.

Contextual variables, mechanism of action, and outcome	Group 1: High engagers, early adopters (easiest fit)	Group 2: High engagers, slow adopters (best fit)	Group 3	
			Low engagers (failed to activate)	Dropouts (failed to meet needs)
Contextual variables				
Preintervention self-efficacy	High (numerous examples of positive behavior change with improved outcomes)	Moderate (some evidence of positive behavior change with variable impact in outcomes)	Low (no evidence of behavior change)	Low (no evidence of behavior change)
Individual identity (includes affect)	New diagnosis	Longstanding diagnosis	Managing T2DM[a] is a struggle and burden	Prospect of managing T2DM competing with psychological issues
	Positive attitude toward life and disease management	Episodic nature of T2DM management leads to frustrations	Described negative emotions (eg, anxiety, depression, anger)	Described negative emotions (eg, anxiety, depression, anger)
	Strong identity that serves as motivation to maintain "healthy" life			
Health beliefs	Proactive, seeks out information	Partial understanding of strategies to manage T2DM	Report barriers to managing T2DM (eg, feelings of deprivation)	Not motivated to better manage T2DM
	Takes ownership	Uncertainty around the impact of certain individual actions	Lack of recognition around proper management	
	Accountable to self			
Support system	Support at home facilitates adherence to diet and recommendations	Well-connected to health care providers for support	No support identified	No support identified
	Project coordinator identified as a source of support			
Competing priorities	None described	None described	Multiple (family, school, work)	Multiple (family, school, work)
Mechanism of action	Performance feedback facilitates self-monitoring	Improved ability to track outcomes increased awareness	Preliminary signs that the app had potential	Participants did not engage with the mobile app
	Data entry reinforces accountability to self	Improved understanding of how individual actions affect T2DM	Mobile app failed to activate mechanisms of change in context	
	Positive outcomes reinforce behavior	Data visualization increased accountability for individual actions		
Outcomes				
HbA_{1c}[b], mean (range)				
Baseline	7.5% (6.2-9.9)	10.0% (8.7-11.1)	8.7% (6.9-10.6)	10.7% (9.7-12.6)
3 months	6.0% (5.2-6.5)	8.3% (7.4-9.2)	8.8% (7.3-10.3)	N/A[c]
PAID5[d], mean (range)				
Baseline	4.3 (1-8)	8.5 (3-14)	10.0 (3-16)	10.0 (5-15)
3 months	3.5 (0-10)	7.3 (2-15)	10.8 (5-14)	N/A

[a]T2DM: type 2 diabetes.

[b]HbA_{1c}: glycated hemoglobin.

[c]N/A: not applicable.

[d]The Problem Areas In Diabetes 5 (PAID5) is a measure of disease-specific self-efficacy that emphasizes well-being. A total score of ≥8 indicates possible diabetes-related emotional distress and warrants further clinical assessment.

Perhaps most notably, individuals in this group demonstrated improvements despite poor disease management at baseline that was significantly worse than group 1 (average HbA_{1c} of 10.0%, refer to Table 2 for complete outcome data):

> *On the odd occasion that my numbers get high, to have the 'let's retest in 3 hours' message [pop up], you know that's been a help as well because it helps me in managing my diabetes.* [SITE C03]

> *That was the key benefit for me. So when I started my A_{1c} was 11.1 and when I got it done, when I saw the nurse, 3 weeks ago, it was 8…A dramatic reduction, and I'm still trying to get it down, but it's a pretty dramatic reduction for me…I don't think I've had an A_{1c} of 8 for a number of years.* [SITE C02]

Similar to the first group, timely feedback and the ability to identify factors that trigger blood glucose spikes increased participants' overall confidence to self-manage their T2DM. Increased confidence also served as a mechanism to resolve unwelcome emotions such as anger and frustration that participants had experienced when struggling to master self-management over the course of their condition:

> *Yeah. No I um, I do really like it, it's kind of kept me going in making me feel a little bit stronger in myself and that with it...Yeah, more confident that I can do it.* [SITE B02]

Group 3: Failure to Activate or Meet Needs—Low Engagers and Dropouts

The final group of C-M-O configurations concerns a group of individuals who had little to no self-efficacy before the intervention; identified a range of psychosocial issues that featured more prominently in their narrative than T2DM; reported a range of competing priorities, including work, family, and school; were uncertain of the benefits of using mobile technology to support self-management; and had a wide range of disease duration (from newly diagnosed to long-standing):

> *It's just a matter of just double checking...And um, if it would connect to the foods I put in to what insulin readings I put in, that would be good, cuz right now it just seems kind of useless. Right now, it's just a matter of double putting in my glucose readings.* [SITE C05]

> *I go back and click on that date and enter all my sugars and meds and what not [all at once]. It's a lot easier than doing it daily—doing it daily it just eats up so much of my time. I only get a half hour lunch break at work usually…I don't want to spend my time fussing with it.* [SITE B05]

Individuals in this group described their experience with managing T2DM as a struggle and viewed the diagnosis as a burden. A range of external barriers were cited that interfered with the ability to self-manage, including a sense of deprivation, unhelpful encounters with health care providers, and a hectic schedule. Participants also described a lack of recognition around proper strategies for self-management and reluctance to engage in basic self-management behaviors (eg, insulin adherence and testing blood sugar levels):

> *I find it hard every time to take my insulin...it's a real chore...Yeah, and trying to find space in my stomach that doesn't hurt...I just feeling like giving up sometimes and not taking it...The times I haven't taken it, it uh, then I get mad at myself which doesn't help the situation.* [SITE C05]

> *It's not really fitting very well because I'm going to be honest I don't really even test my sugars as much as I should. Because sometimes I will miss the time taking my insulin...I haven't tested my sugar in a while.* [SITE A04]

Multiple competing priorities were highlighted, including work and caring for children. Narratives revealed a lack of responsibility for individual actions and a host of cyclical negative emotions, including anxiety, anger, and frustration. Individuals in this group did not identify members of a support system, either from their personal network or their health care team:

> *I don't even know why the clinic was there, it was like, this is a complete waste of my time. You know, I already knew what she told me, like there was no help, you know there was no information offered, I left there empty handed.* [SITE A02]

> *Well the phone was giving me problems at first. So the first thing in my head was 'uh-oh am I going to have a problem with the phone.' And that's when I requested if I could get another phone, but then she said we'll try it again. She went out the room, came back in, I think the first interaction was kinda of–it sucked.* [SITE C01]

Participants highlighted several features of the mobile app that may have represented mechanisms of action but were not activated for this group. The project coordinators at 2 of the sites (who were responsible for introducing the intervention to participants at sites A and C) were perceived to be minimally engaged with participants. This may have represented a missed opportunity for this group of individuals in the absence of strong support networks:

> *I may have gotten an email but there was definitely no phone contact or anything. And I think the only contact that they reached out to me for was making sure I was going to get my [blood glucose test].* [SITE A02]

Overall, individuals in this group failed to integrate the mobile app into routine daily activities, and generally perceived data entry as a burden. Unlike participants in other groups, these individuals viewed the intervention as a duplication of current logbook methods (eg, handwritten) and did not perceive the technological advancement as a relative advantage:

> *Nah. If I forget about it, it sits in my bag, like my pill bag and you go to turn it on and it's dead. Then you gotta plug it in, and then cause you unplug it and put it back in the bag and you forget about it again, right? It's not in my pocket.* [SITE A03]

Failure to activate potential mechanisms of action is particularly relevant, as some participant narratives reveal preliminary signs of intervention potential and positive influence. Nonetheless, the mobile app failed to activate mechanisms of actions for these individuals, and their outcomes remained unchanged (refer to Table 2).

> *And you know [the mobile application] kinda does make you think what you should do differently, and obviously then it's just self-management after that.* [SITE A02]

> *I go back and click on that date and enter all my sugars and meds and what not. It's a lot easier than doing it daily—doing it daily it just eats up so much of my time. Like I only get a half hour lunch break at work usually. I don't want to spend my time fussing with it.* [SITE B05]

Within the first few weeks of use, 3 individuals dropped out of the study and returned their mobile device. Participant narratives revealed that T2DM self-management was competing with prominent psychosocial issues for attention, and was therefore not a high priority. Strong negative emotions were central to each individual's experience, and included feeling overwhelmed with the idea of change, wanting to give up, and struggling to cope. These experiences were compounded by a range of competing priorities, including family, work, and school. As a result, individuals were not motivated to better manage their T2DM:

> *It's overwhelming…I'm not really depressed but I get glum […] Like I get to the point where—the hell with it—and I'll open a can of coke because it's more like a pissed off that I'm going through this, and maybe if I intake enough of bad stuff, I have the seizure or go into the coma or something—not that I'm suicidal or anything…* [SITE B03]

> *I feel deprived of certain things that I want to eat and I know I can't eat it. Uh, it affects me, my mood, some days I'm happy some days I'm sad, um, it I guess that's what triggers my depression in some ways [...] Just sometimes I get frustrated and sometimes feel like giving up.* [SITE C01]

These 3 participants reported limited to no use of the mobile app before dropout, precluding the ability for the intervention to influence change:

> *Even if I did [everything I'm supposed to], starting to use [the app] on a regular basis is gonna be hard too. Because it's not that I'm unwilling which is partly true, I am unwilling—I shouldn't have to do this.* [SITE B03]

Discussion

Principal Findings

The results of this study identify variations in patient characteristics that influenced the adoption and outcome of a mobile-based self-management app to improve self-management in T2DM. To our knowledge, this is the first realist evaluation to systematically link a cluster of patient-level determinants to clinical outcomes with a specified mechanism of action. The results suggest that an individual's self-efficacy, competing priorities, evidence of previous behavior change, and beliefs about Web-based solutions interact to determine the impact on engagement and clinical outcomes. Furthermore, the balance of these characteristics may be useful for identifying individuals who may need more intensive support, informing the allocation of health care resources.

Our findings align with previous qualitative literature identifying increased awareness as a mechanism underlying successful engagement with a Web-based solution for T2DM [43]. Participants whose HbA_{1c} improved >1% reported that the intervention improved their self-efficacy to manage their diabetes, whereas those who failed to achieve these gains reported competing demands that limited engagement with intervention [43]. Self-efficacy is influenced through previous experiences of success, social persuasion and encouragement, social models of success and failure observed from individuals perceived to be similar, and stress and tension [44]. Many participants in this study described the feedback messaging as motivational and encouraging, whereas others reported frustration when glucose readings fell outside the target range and messaging failed to provide encouragement. Feedback messages were triggered in response to available blood glucose data and were not triangulated with other inputs (or lack thereof), which may present an opportunity to further tailor messaging to encourage improvements among poor performers.

Targeting outcome expectations can be easily integrated into a Web-based solution and present one strategy to regulate patient motivation when previous experiences have been unsuccessful [44]. The app included passive access to a resource library that includes a rotating assortment of videos; however, actively directing users to this content may be required for those who require additional support. The current version also included 3 levels of tailored messaging (from a beginner level to more sophisticated content); however, all 3 levels addressed the full range of self-management issues. Interventions designed to promote incremental knowledge gain and experiential and vicarious learning are better positioned to impact individual ability to self-manage [45], suggesting that a graduated approach introducing a few concepts at a time may be more beneficial when implementing complex interventions targeting behavior change. Many individuals with poorly controlled diabetes are not sufficiently confident or motivated to initiate and maintain self-management changes [46], emphasizing the need for mobile self-management apps to explicitly target readiness to change and emphasize increasing self-efficacy to optimize the potential for impact.

A systematic review of mobile apps for diabetes management found 5 of the 6 studies reported positive feedback on usability and feasibility, whereas only 3 reported statistically significant reductions in clinical outcomes such as HbA_{1c} and blood pressure [47]. These heterogeneous findings demonstrate that positive patient feedback does not always accompany clinical improvements, highlighting that a range of factors interact to contribute to a successful impact on outcomes. This underscores the need to identify patient and intervention characteristics that

are likely to facilitate both outcomes. Whether and how a Web-based solution enables or limits the possibility for relationships with professionals, the degree of fit with participants' everyday life and capacity, and pattern visualization of symptoms and feedback are key mechanisms to support successful implementation [48]. Our results also suggest that the impact of Web-based solutions would be enhanced if they were equipped with the ability to adapt to individual users based on the triangulation of available data and proactively identify individuals who require additional support to avoid disengagement.

Among many individuals who failed to achieve engagement and a reduction in HbA_{1c} (group 3), the results provide 2 key insights. First, triangulation with quantitative outcomes (ie, the PAID5) reveals average borderline emotional distress among low engagers and dropouts that warrants further clinical examination. Among low engagers, emotional distress increased slightly from baseline (average PAID5 score of 10.0) to 3 months (average score of 10.8), suggesting it was not effectively addressed as part of the participants' ongoing care. Diabetes-related emotional distress is significantly related to HbA_{1c} levels [49], underscoring the importance of effectively targeting distress to achieve improvements in glycemic control. This could be achieved through more targeted clinical care or by exploring opportunities to address emotional distress as part of a comprehensive Web-based solution. Second, the results reveal an opportunity to modify the Web-based solution or its implementation to address currently unmet needs. Low perceived value and a lack of patient-provider interactions are barriers to engaging in Web-based solutions for T2DM [50,51]. Health literacy is also likely to affect self-management behaviors [52] and is lower among disengaged patients, indicating the need to address underlying cognitive and social skills that determine an individual's motivation and ability to understand and use information to inform healthy behaviors. Similar to our study findings, Lie et al [51] identified that prioritizing other activities and frustrations with the technology led to a loss of motivation among dropouts of a Web-based solution for T2DM. Individuals with T2DM exhibit a variety of dominant coping mechanisms; however, those exhibiting problem-focused and avoidance-focused mechanisms are significantly less likely to be adherent to self-care activities [53]. Web-based solutions can leverage existing measures to evaluate coping strategies to identify individuals who are at risk of suboptimal adherence to inform the provision of additional support(s) as needed.

Individuals manage chronic conditions within different (but not exclusive) nonprofessional contexts where relationships are primarily patterned and unreflexive [54]. Reeves et al [55] demonstrated that health service costs were significantly reduced for individuals who experience greater levels of illness work (eg, crisis prevention and management, symptom management, and disease-specific activities) through their social networks. Illness work was associated with increased self-management, healthy behaviors, and emotional well-being [55], underscoring the value of both harnessing and sustaining the potential of social networks to support the success of self-management interventions. The Web-based solution did not incorporate a social function, nor did the implementation include a mechanism to integrate the solution into the existing social networks. Improvements in clinical outcomes in this study may have been mediated by strong social networks, as these individuals were able to successfully incorporate the solution into the existing routines and negotiate competing priorities. In contrast, individuals who were unsuccessful or disengaged did not identify a pre-existing source of social support. The influence of social mechanisms on individual success should be considered in the design and implementation of Web-based self-management solutions, aligning with the growing recognition that self-management must move beyond an individual-centered view to consider the broader social context [54,56].

Finally, health care provider feedback enhances the impact of mobile self-management apps on HbA_{1c} reduction [22], underscoring the importance of active provider engagement as a support strategy in the early stages of implementation, tailoring support provision throughout the process. In this study, failure to actively engage health care providers during implementation may have contributed to a lack of sustained engagement among those participants for whom the Web-based solution failed to activate change. Further work is needed to understand whether adaptations to the Web-based solution or its implementation would have resulted in different outcomes for these individuals in their contexts. Our findings suggest that targeting outcome expectations, addressing diabetes-related emotional distress, including content to address health literacy, tailoring messaging according to individual coping strategies, and leveraging social networks are worthwhile components to consider as part of a Web-based solution.

Limitations

Participation in the qualitative interviews was voluntary, which introduces the potential of selection bias. To mitigate this, purposive sampling was used to capture the perspectives of participants who had minimal engagement with the Web-based solution. The transferability of the results is limited by the inclusion of a small number of participants across 3 recruitment sites in a confined geographical area. The inclusion of a small number of participants and sites was necessary to achieve a depth of understanding with respect to contextual factors, the features of the Web-based care solution, and how these relate to outcomes. The findings of this study serve as a foundation for future research aimed at achieving a broader understanding of how Web-based solutions work for different patients in a variety of health care contexts. Given the lack of health care provider interaction, it would be beneficial to supplement patient perspectives with those of their health care providers. Finally, a nuanced exploration of the impact of social networks was beyond the scope of this study. Given the strong influence of competing priorities as a contextual factor and the pattern of social support across groups, further work is needed to understand the extent to which both formal and informal social networks play a role in mediating the adoption of self-management behaviors and engagement with the intervention, which may in turn influence clinical outcomes.

Conclusions

An individual's self-efficacy, competing priorities, prior success with behavior change, and beliefs about their health interact to determine engagement with a mobile app to self-manage T2DM and its impact on clinical outcomes. Careful consideration of the balance of these characteristics is likely to help health care providers identify individuals who are more likely to benefit from a Web-based solution and identify those requiring more intensive support and clinical resources. Web-based solutions could also be optimized to support tailored care, including the incorporation of the existing readiness- and risk-assessment measures, to assist in identifying individuals who are at risk of suboptimal adherence to inform the provision of additional support(s) as needed.

Acknowledgments

The authors would like to acknowledge the Ontario Telemedicine Network for their engagement and operational oversight. This study is sponsored by the Ontario Telemedicine Network through funding from Canada Health Infoway and the Ministry of Health and Long-Term Care.

Conflicts of Interest

PA provided consultation services to OTN on projects unrelated to those involved in this study. The remaining authors have no conflicts to declare.

Multimedia Appendix 1

Interview guide.

[PDF File (Adobe PDF File), 37KB - jmir_v20i3e81_app1.pdf]

References

1. World Health Organization. Global Report on Diabetes URL: http://apps.who.int/iris/bitstream/10665/204871/1/9789241565257_eng.pdf [accessed 2017-08-10] [WebCite Cache ID 6scWntmI3]
2. Lipscombe LL, Hux JE. Trends in diabetes prevalence, incidence, and mortality in Ontario, Canada 1995-2005: a population-based study. Lancet 2007 Mar 03;369(9563):750-756. [doi: 10.1016/S0140-6736(07)60361-4] [Medline: 17336651]
3. King H, Aubert RE, Herman WH. Global burden of diabetes, 1995-2025: prevalence, numerical estimates, and projections. Diabetes Care 1998 Sep;21(9):1414-1431. [Medline: 9727886]
4. Wild S, Roglic G, Green A, Sicree R, King H. Global prevalence of diabetes: estimates for the year 2000 and projections for 2030. Diabetes Care 2004 May;27(5):1047-1053. [Medline: 15111519]
5. American Diabetes Association. Diagnosis and classification of diabetes mellitus. Diabetes Care 2014 Jan;37 Suppl 1:S81-S90. [doi: 10.2337/dc14-S081] [Medline: 24357215]
6. Hu FB. Globalization of diabetes: the role of diet, lifestyle, and genes. Diabetes Care 2011 Jun;34(6):1249-1257 [FREE Full text] [doi: 10.2337/dc11-0442] [Medline: 21617109]
7. Katzmarzyk PT, Mason C. Prevalence of class I, II and III obesity in Canada. CMAJ 2006 Jan 17;174(2):156-157 [FREE Full text] [doi: 10.1503/cmaj.050806] [Medline: 16415457]
8. Zimmet P. The burden of type 2 diabetes: are we doing enough? Diabetes Metab 2003;29(4 Pt 2):6S9-618. [Medline: 14502096]
9. Djoussé L, Driver JA, Gaziano JM, Buring JE, Lee IM. Association between modifiable lifestyle factors and residual lifetime risk of diabetes. Nutr Metab Cardiovasc Dis 2013 Jan;23(1):17-22 [FREE Full text] [doi: 10.1016/j.numecd.2011.08.002] [Medline: 21982361]
10. Gaede P, Lund-Andersen H, Parving H, Pedersen O. Effect of a multifactorial intervention on mortality in type 2 diabetes. N Engl J Med 2008 Feb 07;358(6):580-591. [doi: 10.1056/NEJMoa0706245] [Medline: 18256393]
11. Wei M, Gibbons LW, Kampert JB, Nichaman MZ, Blair SN. Low cardiorespiratory fitness and physical inactivity as predictors of mortality in men with type 2 diabetes. Ann Intern Med 2000 Apr 18;132(8):605-611. [Medline: 10766678]
12. Clark M, Hampson SE. Implementing a psychological intervention to improve lifestyle self-management in patients with type 2 diabetes. Patient Educ Couns 2001 Mar;42(3):247-256. [Medline: 11164324]
13. Ridge K, Treasure J, Forbes A, Thomas S, Ismail K. Themes elicited during motivational interviewing to improve glycaemic control in adults with type 1 diabetes mellitus. Diabet Med 2012 Jan;29(1):148-152. [doi: 10.1111/j.1464-5491.2011.03384.x] [Medline: 22082493]
14. Duncan I, Ahmed T, Li QE, Stetson B, Ruggiero L, Burton K, et al. Assessing the value of the diabetes educator. Diabetes Educ 2011;37(5):638-657. [doi: 10.1177/0145721711416256] [Medline: 21878591]
15. Norris SL, Lau J, Smith SJ, Schmid CH, Engelgau MM. Self-management education for adults with type 2 diabetes: a meta-analysis of the effect on glycemic control. Diabetes Care 2002 Jul;25(7):1159-1171. [Medline: 12087014]

16. Davies MJ, Gagliardino JJ, Gray LJ, Khunti K, Mohan V, Hughes R. Real-world factors affecting adherence to insulin therapy in patients with type 1 or type 2 diabetes mellitus: a systematic review. Diabet Med 2013 May;30(5):512-524. [doi: 10.1111/dme.12128] [Medline: 23323988]
17. Raz I, Riddle MC, Rosenstock J, Buse JB, Inzucchi SE, Home PD, et al. Personalized management of hyperglycemia in type 2 diabetes: reflections from a Diabetes Care Editors' Expert Forum. Diabetes Care 2013 Jun;36(6):1779-1788 [FREE Full text] [doi: 10.2337/dc13-0512] [Medline: 23704680]
18. Nam S, Chesla C, Stotts NA, Kroon L, Janson SL. Barriers to diabetes management: patient and provider factors. Diabetes Res Clin Pract 2011 Jul;93(1):1-9. [doi: 10.1016/j.diabres.2011.02.002] [Medline: 21382643]
19. Tang TS, Brown MB, Funnell MM, Anderson RM. Social support, quality of life, and self-care behaviors among African Americans with type 2 diabetes. Diabetes Educ 2008;34(2):266-276. [doi: 10.1177/0145721708315680] [Medline: 18375776]
20. Heisler M, Piette JD. "I help you, and you help me": facilitated telephone peer support among patients with diabetes. Diabetes Educ 2005;31(6):869-879. [doi: 10.1177/0145721705283247] [Medline: 16288094]
21. Holmström IM, Rosenqvist U. Misunderstandings about illness and treatment among patients with type 2 diabetes. J Adv Nurs 2005 Jan;49(2):146-154. [doi: 10.1111/j.1365-2648.2004.03274.x] [Medline: 15641947]
22. Hou C, Carter B, Hewitt J, Francisa T, Mayor S. Do mobile phone applications improve glycemic control (HbA 1c) in the self-management of diabetes? A systematic review, meta-analysis, and GRADE of 14 randomized trials. Diabetes Care 2016 Nov;39(11):2089-2095. [doi: 10.2337/dc16-0346] [Medline: 27926892]
23. 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]
24. Kampmeijer R, Pavlova M, Tambor M, Golinowska S, Groot W. The use of e-health and m-health tools in health promotion and primary prevention among older adults: a systematic literature review. BMC Health Serv Res 2016 Sep 05;16 Suppl 5:290 [FREE Full text] [doi: 10.1186/s12913-016-1522-3] [Medline: 27608677]
25. Garabedian LF, Ross-Degnan D, Wharam JF. Mobile phone and smartphone technologies for diabetes care and self-management. Curr Diab Rep 2015 Dec;15(12):109. [doi: 10.1007/s11892-015-0680-8] [Medline: 26458380]
26. Neumann B, Stefanik M, Gonzalvo J, Weber Z. Diabetes mHealth applications: where are we now? AADE Pract 2016 Aug 16;4(5):28-32. [doi: 10.1177/2325160316661985]
27. Quinn CC, Shardell MD, Terrin ML, Barr EA, Ballew SH, Gruber-Baldini AL. Cluster-randomized trial of a mobile phone personalized behavioral intervention for blood glucose control. Diabetes Care 2011 Sep;34(9):1934-1942 [FREE Full text] [doi: 10.2337/dc11-0366] [Medline: 21788632]
28. Waki K, Fujita H, Uchimura Y, Omae K, Aramaki E, Kato S, et al. DialBetics: a novel smartphone-based self-management support system for type 2 diabetes patients. J Diabetes Sci Technol 2014 Mar 13;8(2):209-215 [FREE Full text] [doi: 10.1177/1932296814526495] [Medline: 24876569]
29. Bell AM, Fonda SJ, Walker MS, Schmidt V, Vigersky RA. Mobile phone-based video messages for diabetes self-care support. J Diabetes Sci Technol 2012 Mar;6(2):310-319 [FREE Full text] [Medline: 22538140]
30. 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;15(4):e86 [FREE Full text] [doi: 10.2196/jmir.2583] [Medline: 23598614]
31. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016 Nov 14;16(1):1158 [FREE Full text] [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]
32. Arnhold M, Quade M, Kirch W. Mobile applications for diabetics: a systematic review and expert-based usability evaluation considering the special requirements of diabetes patients age 50 years or older. J Med Internet Res 2014;16(4):e104 [FREE Full text] [doi: 10.2196/jmir.2968] [Medline: 24718852]
33. Holmen H, Torbjørnsen A, Wahl AK, Jenum AK, Småstuen MC, Arsand E, et al. A mobile health intervention for self-management and lifestyle change for persons with type 2 diabetes, part 2: one-year results from the Norwegian randomized controlled trial RENEWING HEALTH. JMIR mHealth uHealth 2014;2(4):e57 [FREE Full text] [doi: 10.2196/mhealth.3882] [Medline: 25499872]
34. Desveaux L, Agarwal P, Shaw J, Hensel JM, Mukerji G, Onabajo N, et al. A randomized wait-list control trial to evaluate the impact of a mobile application to improve self-management of individuals with type 2 diabetes: a study protocol. BMC Med Inform Decis Mak 2016 Nov 15;16(1):144 [FREE Full text] [doi: 10.1186/s12911-016-0381-5] [Medline: 27842539]
35. Pawson R. The Science of Evaluation: A Realist Manifesto. London, UK: SAGE Publications Ltd; 2013:1-240.
36. Marchal B, van Belle S, van Olmen J, Hoerée T, Kegels G. Is realist evaluation keeping its promise? A review of published empirical studies in the field of health systems research. Evaluation 2012;18(2):192-212. [doi: 10.1177/1356389012442444]
37. McGuire BE, Morrison TG, Hermanns N, Skovlund S, Eldrup E, Gagliardino J, et al. Short-form measures of diabetes-related emotional distress: the Problem Areas in Diabetes Scale (PAID)-5 and PAID-1. Diabetologia 2010 Jan;53(1):66-69. [doi: 10.1007/s00125-009-1559-5] [Medline: 19841892]
38. Schwarzer R, Lippke S, Luszczynska A. Mechanisms of health behavior change in persons with chronic illness or disability: the Health Action Process Approach (HAPA). Rehabil Psychol 2011 Aug;56(3):161-170. [doi: 10.1037/a0024509] [Medline: 21767036]
39. Ontario Telemedicine Network. 2016. URL: https://otn.ca/ [accessed 2018-02-14] [WebCite Cache ID 6xE9z9AIn]

40. Quinn CC, Clough SS, Minor JM, Lender D, Okafor MC, Gruber-Baldini A. WellDoc mobile diabetes management randomized controlled trial: change in clinical and behavioral outcomes and patient and physician satisfaction. Diabetes Technol Ther 2008 Jun;10(3):160-168. [doi: 10.1089/dia.2008.0283] [Medline: 18473689]
41. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706]
42. Kidder LH, Judd CM. Research Methods in Social Relations. 5th edition. New York: Holt, Rinehart, & Winston; 1986.
43. Andrews SM, Sperber NR, Gierisch JM, Danus S, Macy SL, Bosworth HB, et al. Patient perceptions of a comprehensive telemedicine intervention to address persistent poorly controlled diabetes. Patient Prefer Adherence 2017;11:469-478 [FREE Full text] [doi: 10.2147/PPA.S125673] [Medline: 28424543]
44. Bandura A. Self-efficacy. In: Ramachandran VS, editor. Encyclopedia of Human Behavior. New York: Academic Press; 1994:71-81.
45. Wilkinson A, Whitehead L, Ritchie L. Factors influencing the ability to self-manage diabetes for adults living with type 1 or 2 diabetes. Int J Nurs Stud 2014 Jan;51(1):111-122. [doi: 10.1016/j.ijnurstu.2013.01.006] [Medline: 23473390]
46. Nelson KM, McFarland L, Reiber G. Factors influencing disease self-management among veterans with diabetes and poor glycemic control. J Gen Intern Med 2007 Apr;22(4):442-447 [FREE Full text] [doi: 10.1007/s11606-006-0053-8] [Medline: 17372790]
47. Holmen H, Wahl AK, Cvancarova Småstuen M, Ribu L. Tailored communication within mobile apps for diabetes self-management: a systematic review. J Med Internet Res 2017 Jun 23;19(6):e227 [FREE Full text] [doi: 10.2196/jmir.7045] [Medline: 28645890]
48. Vassilev I, Rowsell A, Pope C, Kennedy A, O'Cathain A, Salisbury C, et al. Assessing the implementability of telehealth interventions for self-management support: a realist review. Implement Sci 2015;10:59 [FREE Full text] [doi: 10.1186/s13012-015-0238-9] [Medline: 25906822]
49. Strandberg RB, Graue M, Wentzel-Larsen T, Peyrot M, Rokne B. Relationships of diabetes-specific emotional distress, depression, anxiety, and overall well-being with HbA 1c in adult persons with type 1 diabetes. J Psychosom Res 2014 Sep;77(3):174-179 [FREE Full text] [doi: 10.1016/j.jpsychores.2014.06.015] [Medline: 25149027]
50. Alvarado MM, Kum HC, Gonzalez Coronado K, Foster MJ, Ortega P, Lawley MA. Barriers to remote health interventions for type 2 diabetes: a systematic review and proposed classification scheme. J Med Internet Res 2017 Feb 13;19(2):e28. [doi: 10.2196/jmir.6382] [Medline: 28193598]
51. Lie SS, Karlsen B, Oord ER, Graue M, Oftedal B. Dropout from an eHealth intervention for adults with type 2 diabetes: a qualitative study. J Med Internet Res 2017 May 30;19(5):e187 [FREE Full text] [doi: 10.2196/jmir.7479] [Medline: 28559223]
52. Mackey LM, Doody C, Werner EL, Fullen B. Self-management skills in chronic disease management: what role does health literacy have? Med Decis Making 2016;36(6):741-759. [Medline: 27053527]
53. Albai A, Sima A, Papava I, Roman D, Andor B, Gafencu M. Association between coping mechanisms and adherence to diabetes-related self-care activities: a cross-sectional study. Patient Prefer Adherence 2017;11:1235-1241 [FREE Full text] [doi: 10.2147/PPA.S140146] [Medline: 28761336]
54. Vassilev I, Rogers A, Sanders C, Kennedy A, Blickem C, Protheroe J, et al. Social networks, social capital and chronic illness self-management: a realist review. Chronic Illn 2011 Mar;7(1):60-86. [doi: 10.1177/1742395310383338] [Medline: 20921033]
55. Reeves D, Blickem C, Vassilev I, Brooks H, Kennedy A, Richardson G, et al. The contribution of social networks to the health and self-management of patients with long-term conditions: a longitudinal study. PLoS One 2014;9(6):e98340 [FREE Full text] [doi: 10.1371/journal.pone.0098340] [Medline: 24887107]
56. Gallant MP. The influence of social support on chronic illness self-management: a review and directions for research. Health Educ Behav 2003 Apr;30(2):170-195. [Medline: 12693522]

Abbreviations

C-M-O: context-mechanism-outcome
ITG: immediate treatment group
OTN: Ontario Telemedicine Network
PAID5: Problem Areas In Diabetes 5
T2DM: type 2 diabetes
WLC: wait-list control group

Edited by G Eysenbach; submitted 15.08.17; peer-reviewed by I Vassilev, K Blondon, N Coulson, E Aramaki; comments to author 12.10.17; revised version received 13.11.17; accepted 30.11.17; published 16.03.18.

Please cite as:
Desveaux L, Shaw J, Saragosa M, Soobiah C, Marani H, Hensel J, Agarwal P, Onabajo N, Bhatia RS, Jeffs L
A Mobile App to Improve Self-Management of Individuals With Type 2 Diabetes: Qualitative Realist Evaluation
J Med Internet Res 2018;20(3):e81
URL: http://www.jmir.org/2018/3/e81/
doi:10.2196/jmir.8712
PMID:29549070

Original Paper

Evaluating a Web-Based Social Anxiety Intervention Among University Students: Randomized Controlled Trial

Hugh Cameron McCall[1], BA; Chris G Richardson[2], PhD; Fjola Dogg Helgadottir[3], PhD; Frances S Chen[1], PhD

[1]Department of Psychology, University of British Columbia, Vancouver, BC, Canada

[2]School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada

[3]Vancouver CBT Centre, Vancouver, BC, Canada

Corresponding Author:
Frances S Chen, PhD
Department of Psychology
University of British Columbia
3521-2136 West Mall
Vancouver, BC, V6T 1Z4
Canada
Phone: 1 604 822 2549
Email: frances.chen@psych.ubc.ca

Abstract

Background: Treatment rates for social anxiety, a prevalent and potentially debilitating condition, remain among the lowest of all major mental disorders today. Although computer-delivered interventions are well poised to surmount key barriers to the treatment of social anxiety, most are only marginally effective when delivered as stand-alone treatments. A new, Web-based cognitive behavioral therapy (CBT) intervention called Overcome Social Anxiety was recently created to address the limitations of prior computer-delivered interventions. Users of Overcome Social Anxiety are self-directed through various CBT modules incorporating cognitive restructuring and behavioral experiments. The intervention is personalized to each user's symptoms, and automatic email reminders and time limits are used to encourage adherence.

Objective: The purpose of this study was to conduct a randomized controlled trial to investigate the effectiveness of Overcome Social Anxiety in reducing social anxiety symptoms in a nonclinical sample of university students. As a secondary aim, we also investigated whether Overcome Social Anxiety would increase life satisfaction in this sample.

Methods: Following eligibility screening, participants were randomly assigned to a treatment condition or a wait-list control condition. Only those assigned to the treatment condition were given access to Overcome Social Anxiety; they were asked to complete the program within 4 months. The social interaction anxiety scale (SIAS), the fear of negative evaluation scale (FNE), and the quality of life enjoyment and satisfaction questionnaire—short form (Q-LES-Q-SF) were administered to participants from both conditions during baseline and 4-month follow-up lab visits.

Results: Over the course of the study, participants assigned to the treatment condition experienced a significant reduction in social anxiety (SIAS: $P<.001$, Cohen $d=0.72$; FNE: $P<.001$, Cohen $d=0.82$), whereas those assigned to the control condition did not (SIAS: $P=.13$, Cohen d=0.26; FNE: $P=.40$, Cohen $d=0.14$). Additionally, a direct comparison of the average change in social anxiety in the 2 conditions over the course of the study showed that those assigned to the treatment condition experienced significantly more improvement than those assigned to the control condition (SIAS: $P=.03$, Cohen $d=0.56$; FNE: $P=.001$, Cohen $d=0.97$). Although participants assigned to the treatment condition experienced a slight increase in life satisfaction, as measured by Q-LES-Q-SF scores, and those assigned to the control condition experienced a slight decrease, these changes were not statistically significant (treatment: $P=.35$, Cohen $d=-0.18$; control: $P=.30$, Cohen $d=0.18$).

Conclusions: Our findings indicate that Overcome Social Anxiety is an effective intervention for treating symptoms of social anxiety and that it may have further utility in serving as a model for the development of new interventions. Additionally, our findings provide evidence that contemporary Web-based interventions can be sophisticated enough to benefit users even when delivered as stand-alone treatments, suggesting that further opportunities likely exist for the development of other Web-based mental health interventions.

Trial Registration: ClinicalTrials.gov NCT02792127; https://clinicaltrials.gov/ct2/show/record/NCT02792127 (Archived by WebCite at http://www.webcitation.org/6xGSRh7MG)

(J Med Internet Res 2018;20(3):e91) doi:10.2196/jmir.8630

KEYWORDS

social anxiety; clinical trial; internet; cognitive behavior therapy

Introduction

Background

Social anxiety disorder is one of the most common anxiety disorders, with approximately 13% of people being affected at some point in their lives [1]. Even people who are below the threshold for clinical diagnosis experience substantial distress and functional impairment [2]. Furthermore, research has shown that social anxiety symptoms tend to be persistent at all levels of severity [3]; that social anxiety is closely related to disorders such as substance abuse, disordered eating, and mood disorders [4]; and that the impacts of social anxiety can be severe in both private and professional domains of life when it is left untreated [5-7]. The effectiveness of various psychotherapeutic and pharmaceutical approaches to treating social anxiety—for example, cognitive behavioral therapy (CBT), selective serotonin reuptake inhibitors—is well documented [8,9]; yet, rates of treatment for social anxiety are some of the lowest among all major mental disorders [10], highlighting the need for the development of more accessible treatment options for social anxiety disorder and subclinical social anxiety alike.

Computer-delivered therapy, including computerized CBT, has become increasingly popular in recent years and holds substantial promise for increasing access to effective treatment options for both depression and social anxiety [11]. One of its major advantages lies in its accessibility to individuals who experience geographic, financial, or personal challenges for human-delivered therapy. Notably, because financial and confidentiality concerns are especially common barriers to treatment for people suffering from social anxiety [10], the privacy and relative affordability of computer-delivered therapy may have particular practical utility for treating social anxiety relative to other mental disorders. Moreover, because students infrequently seek help from professionals for mental health–related problems, but tend to be very comfortable with modern digital technologies, computer-delivered therapies may be especially effective among student populations [12]. Given the high rates of anxiety found among young adults [13], and university students in particular [14], exploring the effectiveness of computer-delivered therapies among student populations may be an especially important area of research.

Computer-delivered therapy programs are not a novel innovation; in fact, they date back to the 1960s [15]. However, a meta-analysis found that the effectiveness of Web-based CBT treatment programs that are not supplemented by human-delivered therapy is minimal [16], suggesting that such programs require improvement before they are delivered as stand-alone treatments. A total of 5 common limitations of many Web-based CBT treatments have been identified [15]. First, many treatments do not offer users individualized programs to address their unique symptoms. Second, many programs tend to provide little visual or audio surrogate human contact, despite research attesting to the importance of therapist-client interaction to a program's success [17]. Third, many Web-based CBT treatments lack mechanisms to facilitate adherence, and completion rates of programs can be as low as 1% [18]. Fourth, programs often do not provide corrective feedback to participants who misunderstand important aspects of the CBT process, such as designing behavioral experiments or differentiating between thoughts, emotions, and feelings. Finally, although administration of an appropriate dose of treatment is important to CBT's success (eg, sufficient repetition of CBT exercises) [19], many treatments fail to provide a sufficient dose of treatment to deliver lasting benefits to users.

Overcome Social Anxiety

A Web-based CBT program designed to reduce social anxiety symptoms among stuttering populations, developed specifically to address the aforementioned 5 limitations, has shown promising preliminary results across 3 evaluative studies [20-22]. Originally called CBTpsych, the program has recently been developed into Overcome Social Anxiety, which is no longer tailored specifically toward stuttering populations. Before this study, Overcome Social Anxiety has not received empirical evaluation. The program's clinical content consists of 7 modules, as shown in Textbox 1, which are intended to be completed over a 4- to 6-month period. It was created by 2 professional clinical psychologists, and employs established CBT procedures for treating social anxiety. Although participants in this study were given free access to Overcome Social Anxiety, it is also available to the public for purchase. A screenshot of the program is shown in Multimedia Appendix 1.

Overcome Social Anxiety has built-in mechanisms to address each of the 5 common limitations of Web-based CBT treatments identified above. First, it individualizes treatment programs as a function of participants' responses to questionnaires about the symptoms of social anxiety that they experience. Second, the program provides users with example responses to help ensure that they understand various aspects of the CBT process (eg, how to design effective behavioral experiments), mitigating the need for corrective feedback. Third, to help improve adherence, users are given a limit of 6 months to complete the program and are sent automated email reminders to keep using the program after periods of inactivity. Fourth, the program employs voice recordings of 2 clinical psychologists explaining important aspects of CBT to users, in an effort to more closely mirror psychologist-delivered CBT. Finally, the program is designed to administer a sufficient dose of individualized therapy to effect lasting reductions in users' social anxiety. Specifically, Overcome Social Anxiety employs all aspects of CBT widely accepted today, including the identification of unhelpful thoughts and avoidance behaviors, psychoeducation on emotions and cognitive errors and unhelpful behaviors, the construction of individual models of social anxiety, the employment of cognitive restructuring strategies, and engagement in exposure exercises in the form of behavioral experiments. For more

detailed information about the program, please refer to the studies by Helgadóttir et al [20,21].

Hypotheses

To investigate the effectiveness of Overcome Social Anxiety, we conducted a randomized controlled trial in a population of university students who reported symptoms of social anxiety but had not received a clinical diagnosis for it. Our primary hypothesis was that participants who were given access to Overcome Social Anxiety would experience a greater decrease in social anxiety symptoms over a 4-month period than those assigned to a wait-list control condition.

Textbox 1. Outline of modules employed in Overcome Social Anxiety.

Prequestionnaires

- The questions asked in this section are retrieved from file audit data from cognitive behavioral therapy clinical practice. The user is presented with a list of 37 common social anxiety thoughts (eg, "I can't speak to authority figures") and a list of 26 common avoidance behaviors (eg, verbal presentations). The user ranks how relevant the thoughts and behaviors are to his or her particular symptoms of social anxiety, which the program then uses to individualize the user's course of treatment.

Module 1: Thinking exercises

- The virtual therapeutic relationship is established when the real clinical psychologists introduce themselves via a photograph and a prerecorded sound clip. This section is designed to familiarize the user with the program's methods, such as learning to use feedback via sample answers and voice-overs. Common cognitive errors are described, with exercises designed to educate the user on the relationship between cognition, behaviors, and emotion.

Module 2: Challenging your thinking

- The user is presented with his or her 5 most relevant social anxiety thoughts, and corrective feedback for their particular cognitive errors. The feedback is drawn from a pool of 296 sample answers written for the back end of the program. The user is asked to write 40 different answers to challenge his or her thoughts, using the feedback from the sample answers. In this way, the quality of user responses is shaped across trials.

Module 3: Creating your model

- The user builds his or her own idiosyncratic social anxiety formulation. To prevent errors in constructing the individualized formulation, prewritten symptoms are selected from a list. These include avoided situations, cognitions driving anxiety and avoidance, safety behaviors, mental images, and physical anxiety symptoms. All of the above are presented with detailed education using voice recordings of the clinical psychologists explaining the material.

Module 4: Behavioral experiments

- In this section, the formulation created in the previous section is used to select behavioral experiments to target avoidance and safety behaviors. The user selects an avoided situation from his or her own avoided situations list. The program then creates a behavioral experiment for that situation targeting one or more different cognitions responsible for driving the avoidance and anxiety. The number of experiments to be completed in this section is expected to be around 3-10 for each user. The back end of the program has the potential to create 962 different behavioral experiments for the user. The program determines whether each experiment should be repeated before recommending a novel experiment; this decision is based on whether the user indicates that he or she would still avoid the previously feared situation.

Module 5: Challenge your thinking further

- Advanced cognitive work is presented in this section with a focus on anger. The user is asked to indicate which anger-related beliefs he or she has. The program guides the user through reframing his or her beliefs through a cost-benefit analysis. This is done using material relevant to the user's particular thoughts.

Module 6: Self-processing

- This section targets the maintenance factor of self-focused attention seen in social anxiety. First, to target biased attention in social situations, skills-based attention training [24] is taught to increase the user's control of attention in social situations. Second, rescripting methods are used to help update faulty and unhelpful imagery [25]. Individualization is particularly important in this section, as the user hears a voice-over that rescripts his or her particular image, selected when the user's tailored formulation was constructed.

Module 7: Relapse prevention

- This section deals with relapse prevention and reviews all the former components of the program. Furthermore, as depression is a highly comorbid condition in social anxiety, psychoeducation is focused on preventative behaviors that the user can engage in to maintain treatment gains and reduce negative mood.

Postquestionnaires

- The questionnaires that the user responded to at the beginning of the program are administered again. The program creates histograms to show the user his or her scores before and after treatment. The program then creates a PDF document containing all of the program's materials and the user's individual data for the user to keep, to help maintain the user's treatment gains over time.

Our secondary hypothesis was that those who were given access to the program would experience a more positive change in life satisfaction (ie, a greater increase or a lesser decrease) than those who were not. The inclusion of life satisfaction as a secondary outcome reflects a trend toward the development and use of assessments that complement traditional measures of symptom severity by capturing broader changes in psychosocial functioning and quality of life when assessing the impact of interventions [23].

Methods

Recruitment

A power analysis indicated that we would require a total sample size of 102 participants (ie, 51 per condition) to achieve a power level of 0.80 at the $P<.05$ level of significance, 1-tailed, assuming a moderate effect size of Cohen $d=0.5$. Our final sample consisted of 65 students (female 47/65, 72%) at a large, public Canadian university, who received credit toward undergraduate psychology courses for their participation. Although some research indicates that social anxiety is more prevalent among women than men (eg, [26]), the unequal sex ratio in our sample is likely attributable, at least in part, to the fact that a majority of students in undergraduate psychology courses at the university at which this research was conducted are female. Recruitment was conducted online, through this university's psychology department's "Human Subject Pool" system. The large majority (61/65, 94%) of our participants were first to fourth year undergraduates. Most were either East Asian (31/65, 48%) or white (12/65, 19%), with a mean age of 21.86 years (SD 5.51, range 17-46). All data were collected between September 2016 and June 2017.

All prospective participants completed a prescreening questionnaire online. Participants were deemed eligible for the study if they reported (1) that they were experiencing some degree of social anxiety (see Materials section below), (2) that they had not received treatment for a chronic mental health condition within the 6 months before the commencement of their participation, and (3) that they had not formerly been diagnosed with social anxiety by a clinician. The resulting sample was thus one of students who experienced symptoms of social anxiety but otherwise reported good psychological health at the time of their recruitment. We chose to employ a nonclinical sample for this study to (1) provide a rigorous initial test of Overcome Social Anxiety in a population of participants whom we expected to be less vulnerable than those with clinically diagnosed anxiety, before expanding our testing to a clinical sample, and (2) explore the potential utility of Overcome Social Anxiety in a broad, diverse user base for whom the use of digital technologies is normative and frequent. The full study protocol was approved by the local institutional ethics board.

Materials

Our materials included an eligibility questionnaire, a baseline questionnaire, a follow-up questionnaire, and Overcome Social Anxiety, the program whose effectiveness this study was designed to assess.

The eligibility questionnaire included the 3-item mini-social phobia inventory (Mini-SPIN) [27], as well as 4 social anxiety assessment items from the *Diagnostic and Statistical Manual of Mental Disorders* (5th edition; DSM-5) [28], all of which were selected to assess prospective participants' reliance on "safety behaviors" to quell their social anxiety symptoms. Prospective participants responded (1=not at all, 2=a little bit, 3=somewhat, 4=very much, 5=extremely) to each of the 3 Mini-SPIN items (eg, "I avoid activities in which I am the center of attention") and the 4 DSM-5 items (eg, "I spend a lot of time preparing what to say or how to act in social situations"). Internal consistency for this 7-item scale was good in our sample (Cronbach alpha=.88). Those who endorsed at least 1 of these 7 items with very much or extremely were considered eligible to participate. This criterion was recommended by the third author, who is a professional clinical psychologist. The eligibility questionnaire was also used to screen out those who had formerly been diagnosed with social anxiety disorder by a clinician or had received treatment for a chronic mental health condition within the past 6 months.

The baseline questionnaire consisted of a demographics section, the social interaction anxiety scale (SIAS) [29], the fear of negative evaluation scale (FNE) [30], and the quality of life enjoyment and satisfaction questionnaire—short form (Q-LES-Q-SF) [31]. The multigroup ethnic identity measure—revised [32] was also included in the baseline questionnaire to examine separate research questions about ethnic identity and social anxiety. The follow-up questionnaire consisted of the SIAS, FNE, and Q-LES-Q-SF, with an additional section for treatment condition participants to provide feedback on their experiences using Overcome Social Anxiety.

The FNE is a widely used measure of social anxiety. It consists of 30 binary-choice items, each of which yields a score of 0 or 1, depending on how it is answered, for a total score ranging from 0 to 30. The scale was found to be valid at the time of its initial publication [30], and subsequent research has confirmed that the FNE converges with other measures of social anxiety [33], discriminates between social anxiety and other anxiety disorders [34], and distinguishes between those who exhibit psychological processes characteristic of social anxiety and those who do not [35].

The SIAS measures social anxiety as well. The total of scores for each of its 20 items (0=not at all, 1=slightly, 2=moderately, 3=very, 4=extremely) are summed for a total score between 0 and 80, inclusive. This measure's validity and reliability have been demonstrated [29,36].

Finally, the Q-LES-Q-SF is a measure of life satisfaction. Although it includes 16 items, the last 2 are stand-alone items; therefore, the final score is derived from the sum of responses to the first 14 items (1=very poor, 2=poor, 3=fair, 4=good, 5=very good), each of which addresses satisfaction in a different domain of life (eg, economic status, social relationships). Research has found the Q-LES-Q-SF to be reliable, valid [31], and useful among diverse populations [37].

Procedure

Before recruitment, each participant number (ie, first participant, second participant) was randomly assigned, without stratification, to the treatment or wait-list control condition. This predetermination of condition allowed the attending research assistant to prepare the appropriate study materials in advance of each participant's scheduled lab visit. During their baseline lab visits, those assigned to the control condition were given a brief explanation of the general purpose of the study and, after giving their consent to participate, were asked to complete the baseline questionnaire. Those assigned to the treatment condition completed these same procedures, but were also given a verbal overview of Overcome Social Anxiety, were left alone to browse the program's website for 8 min, and were set up with an Overcome Social Anxiety account by the attending research assistant.

During the 4-month interval between participants' baseline and follow-up lab visits, participants in the treatment condition were sent reminders to continue using the program and information about requirements for the reception of course credit 1 month before the deadline for the completion of certain modules. Participants were granted course credit incrementally depending on their progress in Overcome Social Anxiety. The maximum amount of course credit a participant could receive was redeemable for a 3% grade increase in each of 2 psychology courses. At their follow-up visits, which were scheduled to occur 4 months after the initial visits, participants were asked to complete the follow-up questionnaire. Those assigned to the control condition were then given access to Overcome Social Anxiety. All participants were debriefed before their participation was terminated.

Statistical Analysis

The primary dependent variable was the change in the severity of participants' social anxiety symptoms during the 4 months between their baseline and follow-up lab visits, as measured by the SIAS and the FNE. The secondary dependent variable was the change in participants' life satisfaction over that period of time, as measured by the Q-LES-Q-SF. For each of these 3 measures, within-subjects *t* tests were employed to determine whether baseline scores differed from follow-up scores for each condition. Additionally, a change score was calculated as the difference between each participant's baseline and follow-up score for each measure; between-subjects *t* tests were then employed to compare these change scores, for each measure, across the 2 conditions.

In addition to the aforementioned analyses of data from participants who completed both baseline and follow-up assessments, baseline characteristics of those lost to follow-up were compared with those who completed the study using between-subjects *t* tests and chi-square tests. Multiple imputation using SPSS version 24 (IBM Corp., Armonk, NY) [38] was used to produce 5 datasets with individual missing values on the SIAS, FNE, and Q-LES-Q-SF imputed via a series of multiple linear regression models (monotone method) that predicted missing responses on each outcome measure using sociodemographic and the other outcome variables. The 5 imputed datasets were then used to conduct pooled between-subjects *t* test analyses of difference scores on the SIAS, FNE, and Q-LES-Q-SF between treatment and control conditions and the results compared with the initial between-subjects *t* test analyses based on data from complete cases.

Results

Retention of Participants

Overall, 264 prospective participants completed the eligibility questionnaire, 173 of whom were deemed eligible to participate. Of the 101 students who participated, 51 were assigned to the treatment condition. Out of the participants assigned to the treatment condition, 1 withdrew from the study before receiving the intervention due to a misunderstanding of the time required to participate in the study. A total of 30 (30/50, 60%) treatment condition participants and 35 (35/50, 70%) wait-list control condition participants returned for their follow-up lab visits. Although we have no data that directly address why a substantial proportion of our participants were lost to follow-up, likely reasons are that participants did not require further research participation credit toward undergraduate psychology courses, and that participants—all of whom were university students—were subject to many competing pressures (academic, social, financial, etc), and completing their participation in our study was not a high priority for them.

No significant differences in baseline social anxiety scores (SIAS and FNE), quality of life (Q-LES-Q-SF), age, ethnicity, and gender were found between participants who completed the study and those who did not (all $P>.05$, *t* tests, and chi-square tests). The *P* value for Little test [39] was not significant ($\chi^2_4=3.8$ $P=.43$), providing support for the assumption that the missing follow-up data were missing completely at random. The flow of participants through the trial is displayed in Figure 1.

Baseline Characteristics

The mean SIAS score for the treatment condition was 38.07 (SD 12.75); for the wait-list control condition, the mean SIAS score was 43.60 (SD 13.16). These mean values confirm that our participants had high levels of social anxiety. In comparison, the mean SIAS scores reported by the scale's creators in a general undergraduate sample, and in a community sample, were 19.0 (SD 10.1, n=482) and 18.8 (SD 11.83, n=315), respectively [29]. Indeed, the SIAS means in our sample even exceeded those in a sample of people diagnosed with social phobia (mean 34.6 [SD 16.4], n=243) [29]. Likewise, mean FNE scores for the treatment (mean 21.40 [SD 6.96]) and control conditions (mean 23.37 [SD 5.53]) in our study were high in comparison with the mean scores of undergraduate samples found by the FNE scale's originators (mean 15.47 [SD 8.62], n=205) [30] and others (eg, mean 14.26 [SD 7.72], n=539) [35]. Importantly, independent samples *t* tests revealed no significant differences in mean baseline scores between those assigned to the wait-list control condition and the treatment condition on the SIAS ($t_{63}=1.71$, $P=.09$), FNE ($t_{63}=1.27$, $P=.21$), or Q-LES-Q-SF ($t_{63}=0.35$, $P=.72$). Mean questionnaire scores at baseline, among other statistics, are summarized in Table 1.

Effectiveness of the Intervention

Participants assigned to the treatment condition experienced a significant reduction in social anxiety symptoms according to both the SIAS (t_{29}=3.94, P<.001, Cohen d=0.72) and the FNE (t_{29}=4.48, P<.001, Cohen d=0.82), whereas those assigned to the wait-list control condition did not (SIAS: t_{34}=1.55, P=.13, Cohen d=0.26; FNE: t_{34}=0.85, P=.40, Cohen d=0.14; see Figures 2 and 3). Neither treatment nor control condition participants experienced a significant change in life satisfaction over the course of the study, as measured by the Q-LES-Q-SF (treatment: t_{28}=−0.96, P=.35, Cohen d=−0.18; control: t_{33}=1.05, P=.30, Cohen d=0.18). These results are depicted in Table 2. We also compared the 2 conditions with one another directly through an independent-samples t test using participants' change scores (difference between their baseline and follow-up scores).

Figure 1. Flowchart of participation.

Table 1. Participant characteristics at baseline.

Characteristic	Treatment (n=30)	Control (n=35)	Total (N=65)
Female, n (%)	19 (63)	28 (80)	47 (72)
Ethnicity			
White, n (%)	7 (23)	5 (14)	12 (18)
Asian, n (%)	19 (63)	21 (70)	40 (62)
Other, n (%)	4 (13)	9 (26)	13 (20)
Age in years, mean (SD)	21.53 (4.09)	22.14 (6.53)	21.86 (5.50)
Social interaction anxiety scale, mean (SD)	38.07 (12.75)	43.60 (13.16)	41.05 (13.17)
Fear of negative evaluation scale, mean (SD)	21.46 (6.96)	23.37 (5.53)	22.46 (6.26)
Quality of life enjoyment and satisfaction questionnaire—short form, mean (SD)	0.61 (0.13)	0.62 (0.14)	0.61 (0.14)

Figure 2. Mean social interaction anxiety scale (SIAS) scores at baseline and 4 months. Error bars represent 95% CIs.

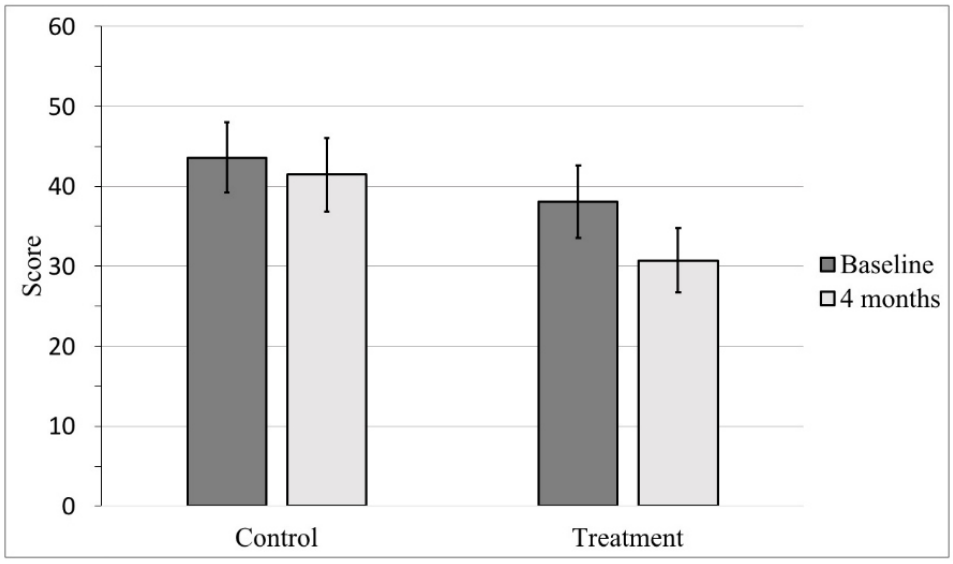


Figure 3. Mean fear of negative evaluation scale (FNE) scores at baseline and 4 months. Error bars represent 95% CI.

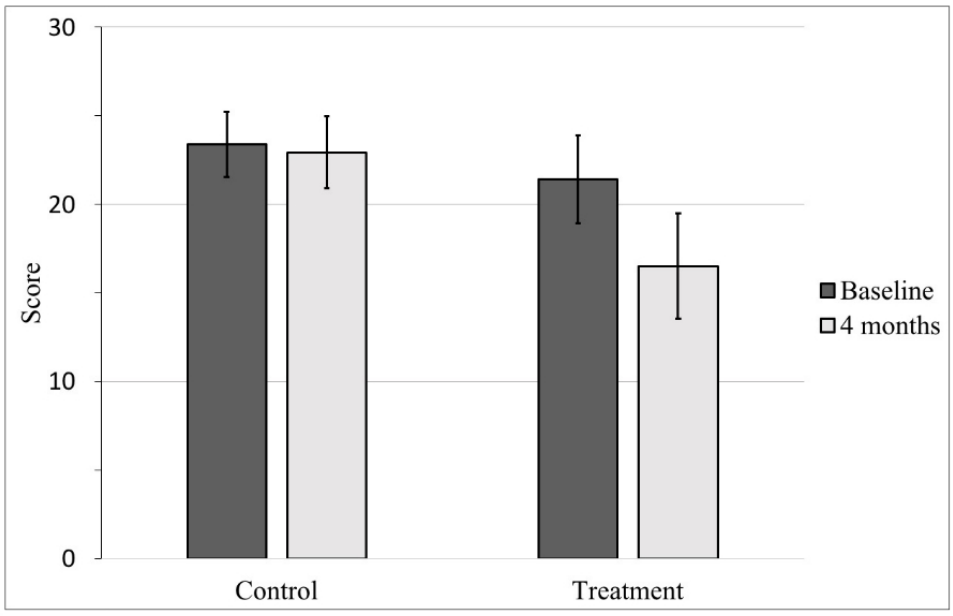


Table 2. Social anxiety symptoms and life satisfaction at baseline and 4-month follow-up.

Condition and time of measurement	Social interaction anxiety scale	Fear of negative evaluation scale	Q-LES-Q-SF[a]
Treatment (n=30)			
Baseline, mean (SD)	38.07 (12.75)	21.40 (6.96)	60.77 (13.32)
4 months, mean (SD)	30.73 (11.12)	16.50 (8.29)	62.81 (15.53)
Change[b]	t_{29}=3.94, *P*<.001, *d*=0.72	t_{29}=4.48, *P*<.001, *d*=0.82	t_{28}=−0.96, *P*=.35, *d*=−0.18
Control (n=35)			
Baseline, mean (SD)	43.60 (13.16)	23.37 (5.53)	61.99 (14.17)
4 months, mean (SD)	41.43 (13.82)	22.91 (6.13)	60.66 (13.77)
Change[b]	t_{34}=1.55, *P*=.13, *d*=0.26	t_{34}=0.85, *P*=.40, *d*=0.14	t_{33}=1.05, *P*=.30, *d*=0.18

[a] Q-LES-Q-SF: Quality of life enjoyment and satisfaction questionnaire—short form. In each condition, 1 participant did not complete the Q-LES-Q-SF during the follow-up visits. Data presented in this column for 4-month follow-up and 4-month change are thus based on responses from 29 treatment condition and 34 control condition participants.

[b]This row displays the results of within-subjects *t* tests comparing participants' baseline scores with their follow-up scores for each measure.

Those assigned to the treatment condition experienced a significantly greater reduction in anxiety than those assigned to the control condition, for both the SIAS (t_{63}=2.25, *P*=.03, Cohen *d*=0.56) and FNE (equal variances not assumed; $t_{42.54}$=3.65, *P*=.001, Cohen *d*=0.97). No significant differences were found between the 2 conditions' changes in Q-LES-Q-SF scores (t_{61}=−1.41, *P*=.16, Cohen *d*=0.36).

To assess the potential impact of missing data, multiple imputation was then used to create 5 imputed datasets, and independent-samples *t* tests were used to compare participants' SIAS, FNE, and Q-LES-Q-SF change scores across treatment and control conditions. Those assigned to the treatment condition experienced a significantly greater reduction in anxiety than those assigned to the control condition, for both the SIAS (t_{172}=2.12, *P*=.04, Cohen *d*=0.32) and FNE (t_{110}=3.63, *P*<.001, Cohen *d*=0.69). No significant differences were found between the 2 conditions' changes in Q-LES-Q-SF scores (t_{48}=−1.01, *P*=.32, Cohen *d*=0.29).

Acceptability of the Intervention

Feedback obtained from the 30 treatment condition participants at follow-up was generally positive. Overall, 80% (24/30) of users reported that the quality of support they received from Overcome Social Anxiety was "good" or "excellent," 87% (26/30) reported "generally" or "definitely" getting the kind of support they wanted from the program, 83% (25/30) responded with "I think so" or "definitely" when asked whether they would recommend Overcome Social Anxiety to a friend in need of similar help, 80% (24/30) reported being "mostly satisfied" or "very satisfied" with the program overall, and 77% (23/30) reported that the program "helped" them or "helped [them] a great deal" to deal more effectively with their problems.

Discussion

Principal Findings

The primary dependent variable of this study was change in social anxiety symptoms over a 4-month period. Our hypothesis, namely, that those assigned to the treatment condition would experience a greater reduction in social anxiety than those assigned to the wait-list control condition, was well supported by both the SIAS and FNE results. Indeed, the effect sizes for treatment condition participants' reduction in social anxiety symptoms over the course of the study, as measured by both the SIAS (Cohen *d*=0.72) and FNE (Cohen *d*=0.82), were approximately *triple* the mean effect size of 6 stand-alone, internet-based CBT treatments for anxiety and depression (Cohen *d*=0.24) found in a meta-analysis [16]. A direct comparison of the treatment and wait-list control conditions' 4-month change scores on the FNE also revealed a large effect size (Cohen *d*=0.97). In fact, this effect was larger than the mean effect size, calculated between conditions, of 19 randomized controlled trials of computer-aided interventions for anxiety disorders found in a review (mean Cohen *d*=0.96) [40], despite the fact that the interventions in this review all benefited from therapist support. This is surprising, given the clear relationship the authors of this review found between the effect sizes in these 19 studies and the amount of therapist support those studies' participants received. In other words, even though therapist support appears to contribute substantially to the effectiveness of computer-delivered CBT for anxiety, our findings indicated that Overcome Social Anxiety is comparably effective to therapist-assisted interventions when delivered as a stand-alone treatment.

Our secondary hypothesis was that those assigned to the treatment condition would experience a more positive change in life satisfaction from baseline to follow-up than those assigned to the control condition. Our results did not support this hypothesis, as neither the differences between treatment condition participants' baseline and follow-up Q-LES-Q-SF scores (*P*=.30, Cohen *d*=−0.18) nor the differences between treatment and control participants' baseline-to-follow-up change scores on that measure (*P*=.16, Cohen *d*=0.36) were significant. Descriptively, however, participants in the treatment condition showed a small increase in life satisfaction over the 4-month period, whereas participants in the control condition showed a small decrease. It may be that further gains in life satisfaction require additional time, post treatment, to accrue. Alternately, the most clear-cut effects of the intervention may be relatively

specific to social anxiety symptoms, with less generalization to life satisfaction.

Finally, participants' extreme scores on both the SIAS and the FNE suggest that—although we excluded those who had formerly been diagnosed with social anxiety—many of our participants may have met diagnostic criteria for the disorder had they previously received help. To the extent that this is true, it provides an unintended illustration of the severity of social anxiety's undertreatment among university students.

Limitations and Future Research

On account of a smaller initial sample size and higher dropout rate than we had anticipated, we failed to reach our target sample size. The lack of significant differences in baseline characteristics between participants who were lost to follow-up and those who completed the study and the results of the multiple imputation support the interpretation of results based on the analyses of complete cases. However, our sample's unequal sex distribution, generally homogenous age distribution, and unanimously high level of education all limit the generalizability of our findings to other populations. Although not presented in the results, separate subgroup analyses of males and females showed a similar pattern of results to the pooled results, suggesting that the intervention is equally effective in males and females.

In terms of future research, it would be informative to investigate the program's effectiveness among clinical populations and populations less comfortable with digital technology than the students who participated in our study.

Because the wait-list control condition did not receive any treatment, the treatment condition's greater reduction in social anxiety symptoms may be attributable, at least in part, to an expectancy effect. It should be noted, however, that because Overcome Social Anxiety is targeted primarily toward those for whom access to human-delivered treatment is limited, any benefit to users—whether attributable to the content of the program itself or attributable to an expectancy effect—is a desirable outcome. Nonetheless, comparing the effectiveness of Overcome Social Anxiety with that of other Web-based and human-delivered treatments would be an important avenue for future research.

Additionally, the research assistants who provided instructions to participants during the follow-up session were aware of participants' conditions due to prior correspondence and interaction with them. This introduces the possibility of an experimenter effect in our outcome measures, although we believe that any such effect would have been mitigated by the fact that research assistants were positioned such that they were unable to see the computer screen on which participants completed questionnaires. However, future research evaluating the effectiveness of Overcome Social Anxiety would benefit from blinding those who administer outcome measures to each participant's condition.

Interestingly, the fact that participation in this study was extrinsically motivated through our granting of course credit may have affected our findings in 2 opposing ways. On one hand, it is reasonable to suppose that the program may have been less effective among our extrinsically motivated participants than it would be among the program's intended users, whom we presume to have predominantly intrinsic motivations for using it; for example, extrinsically motivated users may simply put less effort than intrinsically motivated users into maximizing their benefit from the program. On the other hand, our incremental granting of course credit—which was contingent upon the number of modules each treatment condition participant completed—may have artificially inflated adherence in our sample. Thus, participants in our sample may have benefited less from their usage of the program, per hour of use or per module completed, but spent more hours and completed more modules overall than the typical intended user.

Although our findings provide preliminary evidence that Overcome Social Anxiety is effective, further research will be required to elucidate which elements of the program contributed most to its effectiveness. Future research comparing several variants of Web-based interventions could prove particularly fruitful in discerning which aspects of a Web-based intervention are most important to its ultimate success.

Conclusions

Our findings show that Overcome Social Anxiety is an effective program for reducing social anxiety symptoms. Additionally, the mechanisms it employs to overcome the limitations of previous Web-based CBT interventions appear to have been successful and may thus help guide the development of future Web-based treatments. Finally, our results indicate that developments at the intersection of psychology and technology are now sufficient to create effective, stand-alone, computer-delivered therapy programs, highlighting the opportunity for further research in this exciting area.

Acknowledgments

The authors would like to thank Keith Patena and Ashley Clark for their invaluable assistance in organizing and conducting this study. Additionally, the authors would like to thank all the participants, particularly those who were ineligible to receive course credit for their participation at follow-up but returned nonetheless out of goodwill.

Conflicts of Interest

The third author, FH, is a director of AI-Therapy, the company that created Overcome Social Anxiety. FH, as a registered psychologist, provided critical input for the design of the study. She neither provided funding for the study nor had access to the data. All other authors declare no conflicts of interest.

Multimedia Appendix 1

Screenshot of Overcome Social Anxiety.

[PNG File, 866KB - jmir_v20i3e91_app1.png]

Multimedia Appendix 2

CONSORT-EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 530KB - jmir_v20i3e91_app2.pdf]

References

1. Kessler RC, Petukhova M, Sampson NA, Zaslavsky AM, Wittchen HU. Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. Int J Methods Psychiatr Res 2012 Sep;21(3):169-184 [FREE Full text] [doi: 10.1002/mpr.1359] [Medline: 22865617]
2. Dell'Osso L, Rucci P, Ducci F, Ciapparelli A, Vivarelli L, Carlini M, et al. Social anxiety spectrum. Eur Arch Psychiatry Clin Neurosci 2003 Dec;253(6):286-291. [doi: 10.1007/s00406-003-0442-5] [Medline: 14714117]
3. Reich J, Goldenberg I, Vasile R, Goisman R, Keller M. A prospective follow-along study of the course of social phobia. Psychiatry Res 1994 Dec;54(3):249-258. [Medline: 7792329]
4. Antony MM, Federici A, Stein MB. Overview and introduction to anxiety disorders. In: Oxford Handbook of Anxiety and Related Disorders. New York, NY, US: Oxford University Press; 2009:3-15.
5. Davidson JR, Hughes DL, George LK, Blazer DG. The epidemiology of social phobia: findings from the Duke Epidemiological Catchment Area Study. Psychol Med 1993 Aug;23(3):709-718. [Medline: 8234577]
6. Schneier FR, Heckelman LR, Garfinkel R, Campeas R, Fallon BA, Gitow A, et al. Functional impairment in social phobia. J Clin Psychiatry 1994 Aug;55(8):322-331. [Medline: 8071299]
7. Stein MB, Torgrud LJ, Walker JR. Social phobia symptoms, subtypes, and severity: findings from a community survey. Arch Gen Psychiatry 2000 Nov;57(11):1046-1052. [Medline: 11074870]
8. Gould RA, Buckminster S, Pollack MH, Otto MW, Massachusetts LY. Cognitive behavioral and pharmacological treatment for social phobia: a meta-analysis. Clinical Psychology: Science and Practice 1997;4(4):291-306. [doi: 10.1111/j.1468-2850.1997.tb00123.x]
9. Rodebaugh TL, Holaway RM, Heimberg RG. The treatment of social anxiety disorder. Clin Psychol Rev 2004 Nov;24(7):883-908. [doi: 10.1016/j.cpr.2004.07.007] [Medline: 15501560]
10. Olfson M, Guardino M, Struening E, Schneier FR, Hellman F, Klein DF. Barriers to the treatment of social anxiety. Am J Psychiatry 2000 Apr;157(4):521-527. [Medline: 10739410]
11. Kaltenthaler E, Cavanagh K. Computerised cognitive behavioural therapy and its uses. Prog Neurol Psychiatry 2010 Jun 09;14(3):22-29. [doi: 10.1002/pnp.163]
12. Davies EB, Morriss R, Glazebrook C. Computer-delivered and web-based interventions to improve depression, anxiety, and psychological well-being of university students: a systematic review and meta-analysis. J Med Internet Res 2014;16(5):e130 [FREE Full text] [doi: 10.2196/jmir.3142] [Medline: 24836465]
13. Collins ME, Mowbray CT. Higher education and psychiatric disabilities: national survey of campus disability services. Am J Orthopsychiatry 2005 Apr;75(2):304-315. [doi: 10.1037/0002-9432.75.2.304] [Medline: 15839766]
14. Bayram N, Bilgel N. The prevalence and socio-demographic correlations of depression, anxiety and stress among a group of university students. Soc Psychiatry Psychiatr Epidemiol 2008 Aug;43(8):667-672. [doi: 10.1007/s00127-008-0345-x] [Medline: 18398558]
15. Helgadóttir FD, Menzies RG, Onslow M, Packman A, O'Brian S. Online CBT I: bridging the gap between Eliza and modern online CBT treatment packages. Behav Change 2009 Dec 01;26(4):245-253. [doi: 10.1375/bech.26.4.245]
16. 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]
17. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
18. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
19. Craske MG, Roy-Byrne P, Stein MB, Sullivan G, Hazlett-Stevens H, Bystritsky A, et al. CBT intensity and outcome for panic disorder in a primary care setting. Behav Ther 2006 Jun;37(2):112-119. [doi: 10.1016/j.beth.2005.05.003] [Medline: 16942966]
20. Helgadóttir FD, Menzies RG, Onslow M, Packman A, O'Brian S. Online CBT II: a phase I trial of a standalone, online CBT treatment program for social anxiety in stuttering. Behav Change 2009;26(4):254-270. [doi: 10.1375/bech.26.4.254]

21. Helgadóttir FD, Menzies RG, Onslow M, Packman A, O'Brian S. A standalone Internet cognitive behavior therapy treatment for social anxiety in adults who stutter: CBTpsych. J Fluency Disord 2014 Sep;41:47-54. [doi: 10.1016/j.jfludis.2014.04.001] [Medline: 25173456]
22. Menzies R, O'Brian S, Lowe R, Packman A, Onslow M. International phase II clinical trial of CBTPsych: a standalone internet social anxiety treatment for adults who stutter. J Fluency Disord 2016 Jun;48:35-43. [doi: 10.1016/j.jfludis.2016.06.002] [Medline: 27498893]
23. Barrera TL, Norton PJ. Quality of life impairment in generalized anxiety disorder, social phobia, and panic disorder. J Anxiety Disord 2009 Dec;23(8):1086-1090 [FREE Full text] [doi: 10.1016/j.janxdis.2009.07.011] [Medline: 19640675]
24. Wells A. The attention training technique: theory, effects, and a metacognitive hypothesis on auditory hallucinations. Cogn Behav Pract 2007 May;14(2):134-138. [doi: 10.1016/j.cbpra.2006.01.010]
25. Holmes EA, Arntz A, Smucker MR. Imagery rescripting in cognitive behaviour therapy: images, treatment techniques and outcomes. J Behav Ther Exp Psychiatry 2007 Dec;38(4):297-305. [doi: 10.1016/j.jbtep.2007.10.007] [Medline: 18035331]
26. Caballo VE, Salazar IC, Irurtia MJ, Arias B, Hofmann SG. Measuring social anxiety in 11 countries. Eur J Psychol Assess 2010 Jan;26(2):95-107. [doi: 10.1027/1015-5759/a000014]
27. Connor KM, Kobak KA, Churchill LE, Katzelnick D, Davidson JR. Mini-SPIN: a brief screening assessment for generalized social anxiety disorder. Depress Anxiety 2001;14(2):137-140. [Medline: 11668666]
28. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (5th ed). Arlington, VA, US: American Psychiatric Publishing; 2013.
29. Mattick RP, Clarke JC. Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. Behav Res Ther 1998 Apr;36(4):455-470. [Medline: 9670605]
30. Watson D, Friend R. Measurement of social-evaluative anxiety. J Consult Clin Psychol 1969 Aug;33(4):448-457. [Medline: 5810590]
31. Endicott J, Nee J, Harrison W, Blumenthal R. Quality of life enjoyment and satisfaction questionnaire: a new measure. Psychopharmacol Bull 1993;29(2):321-326. [Medline: 8290681]
32. Phinney JS, Ong AD. Conceptualization and measurement of ethnic identity: current status and future directions. J Couns Psychol 2007;54(3):271-281. [doi: 10.1037/0022-0167.54.3.271]
33. Heimberg RG, Mueller GP, Holt CS, Hope DA, Liebowitz MR. Assessment of anxiety in social interaction and being observed by others: the social interaction anxiety scale and the social phobia scale. Behav Ther 1992;23(1):53-73. [doi: 10.1016/S0005-7894(05)80308-9]
34. Stopa L, Clark DM. Social phobia and interpretation of social events. Behav Res Ther 2000 Mar;38(3):273-283. [Medline: 10665160]
35. Stopa L, Clark DM. Social phobia: comments on the viability and validity of an analogue research strategy and British norms for the Fear of Negative Evaluation Questionnaire. Behav Cogn Psychother 2001 Oct;29(04):423-430. [doi: 10.1017/S1352465801004039]
36. Brown EJ, Turovsky J, Heimberg RG, Juster HR, Brown TA, Barlow DH. Validation of the Social Interaction Anxiety Scale and the Social Phobia Scale across the anxiety disorders. Psychol Assess 1997;9(1):21-27. [doi: 10.1037/1040-3590.9.1.21]
37. Stevanovic D. Quality of Life Enjoyment and Satisfaction Questionnaire-short form for quality of life assessments in clinical practice: a psychometric study. J Psychiatr Ment Health Nurs 2011 Oct;18(8):744-750. [doi: 10.1111/j.1365-2850.2011.01735.x] [Medline: 21896118]
38. IBM Corp. Armonk, NY; 2016. IBM SPSS Statistics 24.0 URL: http://www-01.ibm.com/support/docview.wss?uid=swg27047057 [WebCite Cache ID 6xPlsg2nr]
39. Little RJ. A test of missing completely at random for multivariate data with missing values. J Am Stat Assoc 1988;83(404):1198-1202. [doi: 10.1080/01621459.1988.10478722]
40. 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]

Abbreviations

CBT: cognitive behavioral therapy
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, 5th edition
FNE: fear of negative evaluation scale
Mini-SPIN: mini-social phobia inventory
Q-LES-Q-SF: quality of life enjoyment and satisfaction questionnaire—short form
SIAS: social interaction anxiety scale

Edited by G Eysenbach; submitted 09.08.17; peer-reviewed by EB Davies, K Reynolds; comments to author 06.12.17; revised version received 20.12.17; accepted 22.12.17; published 21.03.18.

Please cite as:
McCall HC, Richardson CG, Helgadottir FD, Chen FS
Evaluating a Web-Based Social Anxiety Intervention Among University Students: Randomized Controlled Trial
J Med Internet Res 2018;20(3):e91
URL: http://www.jmir.org/2018/3/e91/
doi:10.2196/jmir.8630
PMID:29563078

Review

Just-in-Time Feedback in Diet and Physical Activity Interventions: Systematic Review and Practical Design Framework

Susan M Schembre[1*], RD, MS, PhD; Yue Liao[1*], MPH, PhD; Michael C Robertson[1], MPH; Genevieve Fridlund Dunton[2], MPH, PhD; Jacqueline Kerr[3], BA, MSc, PhD; Meghan E Haffey[4], BA; Taylor Burnett[5], BS; Karen Basen-Engquist[1], MPH, PhD; Rachel S Hicklen[6], BA

[1]Department of Behavioral Science, Division of Cancer Control and Population Sciences, The University of Texas MD Anderson Cancer Center, Houston, TX, United States

[2]Institute for Health Promotion & Disease Prevention, Department of Preventive Medicine, University of Southern California, Los Angeles, CA, United States

[3]Division of Behavioral Medicine, Department of Family Medicine and Public Health, University of California, San Diego, San Diego, CA, United States

[4]Department of Epidemiology, University of Texas School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, United States

[5]Department of Family and Consumer Sciences, College of Health Science, Sam Houston State University, Huntsville, TX, United States

[6]Research Medical Library, The University of Texas MD Anderson Cancer Center, Houston, TX, United States

[*]these authors contributed equally

Corresponding Author:
Susan M Schembre, RD, MS, PhD
Department of Behavioral Science
Division of Cancer Control and Population Sciences
The University of Texas MD Anderson Cancer Center
Unit 1330
PO Box 301439
Houston, TX, 77230-1439
United States
Phone: 1 713 563 5858
Fax: 1 713 745 4468
Email: sschembre@mdanderson.org

Abstract

Background: The integration of body-worn sensors with mobile devices presents a tremendous opportunity to improve just-in-time behavioral interventions by enhancing bidirectional communication between investigators and their participants. This approach can be used to deliver supportive feedback at critical moments to optimize the attainment of health behavior goals.

Objective: The goals of this systematic review were to summarize data on the content characteristics of feedback messaging used in diet and physical activity (PA) interventions and to develop a practical framework for designing just-in-time feedback for behavioral interventions.

Methods: Interventions that included just-in-time feedback on PA, sedentary behavior, or dietary intake were eligible for inclusion. Feedback content and efficacy data were synthesized descriptively.

Results: The review included 31 studies (15/31, 48%, targeting PA or sedentary behavior only; 13/31, 42%, targeting diet and PA; and 3/31, 10%, targeting diet only). All studies used just-in-time feedback, 30 (97%, 30/31) used personalized feedback, and 24 (78%, 24/31) used goal-oriented feedback, but only 5 (16%, 5/31) used actionable feedback. Of the 9 studies that tested the efficacy of providing feedback to promote behavior change, 4 reported significant improvements in health behavior. In 3 of these 4 studies, feedback was continuously available, goal-oriented, or actionable.

Conclusions: Feedback that was continuously available, personalized, and actionable relative to a known behavioral objective was prominent in intervention studies with significant behavior change outcomes. Future research should determine whether all or some of these characteristics are needed to optimize the effect of feedback in just-in-time interventions.

(J Med Internet Res 2018;20(3):e106) doi:10.2196/jmir.8701

KEYWORDS

health behavior; diet; exercise; task performance and analysis; Internet; mHealth; accelerometer; activity monitor; self-tracking; wearable sensors

Introduction

Recent advancements in technology, particularly the advent of activity monitors and other wearable body sensors, have the potential to influence innovations in diet and physical activity (PA) assessment and interventions. According to 2013 Pew statistics [1], 7 in 10 US adults report tracking at least one health indicator (eg, weight, diet, exercise, blood pressure, blood sugar, or sleep patterns), and of people who track a health indicator, 46% report that such tracking has changed their approach to health maintenance, 40% say it has led them to engage with health care providers, and 34% say it has affected a decision on how to treat an illness or condition. However, only 21% of people who track a health indicator say that they used some form of technology to assist their tracking efforts. These data highlight that health-tracking technology is surprisingly underutilized as a resource to motivate health behavior change. However, studies testing the efficacy of health-tracking technology to motivate behavior change alone or as part of a theory-based behavioral intervention have shown that such technology-based approaches produce null to modest short-term improvements in health behaviors or weight loss compared with traditional approaches (eg, in-person coaching and telephone coaching) [2-4]. One potential reason for these lackluster findings is that our current behavioral theories have not yet been adapted to leverage the advantages of health-tracking technologies.

Wearable sensors, particularly Internet-connected sensors, can dramatically enrich the temporality and frequency of health behavior data collection by facilitating self-monitoring and reducing self-report biases. Another important but less realized advantage of wearable sensor technology is its bidirectional communication capability. The latest health trackers are equipped with interactive software apps housing algorithms that allow data to be processed in real time to deliver actionable feedback at critical moments in a person's daily life to facilitate the attainment of predetermined health behavior goals. These features are likely to enhance bidirectional communication between investigators and their study participants or between patients and their health care providers, thereby improving the users' engagement with the technology and subsequently facilitating intervention adherence and improving health outcomes. Technology-enhanced interventions are likely the future of behavior change research; however, the use of fast-advancing technologies that enable just-in-time interventions is outpacing the adaptation of theory-based intervention design [5].

Performance feedback is a key, theory-based behavior change strategy [5] that has not been optimally adapted for technology-enhanced interventions. Performance feedback is historically defined as actions taken by (an) external agent(s) to provide information regarding some aspect(s) of one's task-specific performance [6]. Several behavior change theories, including control theory [7,8], goal-setting theory [9], and social cognitive theory (SCT) [10], deem feedback to be an important component of successful behavior change. Although the rationale for using feedback varies across these theories, each characterizes feedback as a self-regulation strategy that reveals to people their progress in relation to their goal(s) [7-10]. In addition, the feedback intervention theory (FIT) posits that feedback, as a component of behavioral interventions, motivates behavior change by focusing one's attention on the behavioral task itself [6], which introduces the importance of timing in the delivery of feedback messages. Collectively, these theories indicate that feedback should be personalized and goal-oriented and presented when attention could be refocused to improve the likelihood of goal attainment. These characteristics are consistent with the strategy of using feedback in just-in-time behavior change interventions [5,11,12]. Just-in-time interventions are those which are delivered when there is an opportunity for positive change. Performance feedback represents a type of support that can be delivered at important decision points relative to a specific behavior and the attainment of related behavioral goals. However, the use of feedback and its content varies widely within behavioral domains, such as learning, professional care practice, and employee performance [13]. Furthermore, there is limited systematic analysis of the characteristics and use of feedback in health behavior interventions. The paucity of literature in the area of health behavior change might be limiting our ability to optimize feedback content to obtain the greatest intervention effect.

Several studies have reviewed the efficacy of digital health technology to promote weight control, PA, and healthy diets. These reviews generally support the use of technology for self-monitoring and intervention delivery, but they also acknowledge that the content and design of future interventions will need more rigorous evaluation to optimize their effects [12,14-16]. In a historical perspective and meta-analysis, Kluger and Denisi [6] found that feedback interventions significantly improved performance. The authors concluded that feedback that is more likely to have a positive effect on behavior change is specific to a familiar task (ie, personalized) and attracts attention to discrepancies between current performance and a desired goal or target at the task level (ie, goal-oriented). They further suggested that feedback should contain cues that support one's performance of the task (ie, should be action-oriented). In a more recent systematic review and meta-analysis, Sherrington and colleagues [17] showed that participants enrolled in weight loss interventions that provided personalized feedback lost on average 2.13 kg (P<.001) more weight than those in control groups who received no feedback. Although their review provided evidence that feedback was an effective behavior change strategy in weight loss interventions, it lacked an in-depth characterization of effective feedback components to be applied in future studies. Collectively, these reviews

support the use of feedback to motivate health behavior change, particularly in the context of diet and PA, and point to key characteristics of potentially effective feedback. To date, however, there has been no framework developed for designing feedback to be used technology-enhanced behavioral interventions.

The primary goal of this review was to (1) Provide a review of diet and physical activity (PA) interventions that use just-in-time feedback as a behavior change technique (BCT); (2) Characterize key aspects of the reviewed studies' feedback content characteristics, prompting style, and delivery methods; and (3) describe how the implementation of these key aspects differed by studies that found significant effects of using feedback to motivate behavior change (intervention efficacy). Our secondary goal was to develop a practical framework for designing feedback that could be incorporated into technology-enhanced just-in-time interventions. We focused on feedback characteristics inferred from the theoretical and historical foundations of the use of performance feedback as a health behavior change strategy: timeliness, personalization, action orientation, and goal orientation.

Methods

Literature Search

Two authors (SMS and YL) with the assistance of a medical librarian (RSH) devised systematic strategies to search the Ovid MEDLINE, Ovid EMBASE, PubMed, Cochrane Library, Scopus, and PsychInfo databases for all relevant literature published through December 2016. Searches were limited articles written in the English language and conducted in humans. Database search strategies included the use of controlled vocabulary (eg, Medical Subject Headings and Emtree) and keywords to identify studies addressing PA or diet in conjunction with feedback. Keywords included physical activity, exercise, diet, eating, intervention, and feedback. Additionally, the bibliographies of topically relevant review papers and all included studies were examined to identify any additional studies.

Study Inclusion and Exclusion Criteria

Eligible studies included (1) Just-in-time feedback as an intervention component and (2) Targeted behavior changes that included PA, sedentary behavior, or dietary intake. Just-in-time feedback was defined on a case-by-case basis as any feedback that focused on participants' daily PA, sedentary behavior, or dietary intake and that was provided within 1 min to 1 day of assessing current performance, as appropriate to each intervention or behavior change goal. Studies with multiple feedback components were included, but at least one type of delivered feedback must have met this definition of just-in-time feedback.

Studies were excluded if (1) the intervention-targeted behaviors were off-topic (eg, studies of clinical education, personnel, management, medication adherence, blood glucose self-monitoring, and symptom management); (2) no intervention outcome results were reported (eg, protocol papers); (3) the time frame for providing feedback was greater than 1 day (eg, weekly performance summaries); (4) no or inadequate feedback was provided (eg, studies of performance tracking without evaluation); or (5) they reported the secondary outcomes of an included intervention.

Data Extraction and Management

Data were extracted into a structured coding form according to Preferred Reporting of Systematic Reviews and Meta-Analyses (PRISMA) guidelines [18] and the Cochrane Handbook for Systematic Reviews of Interventions [19]. Five reviewers (SMS, YL, MCR, MEH, and TB) extracted characteristics and outcomes from all studies using a standardized data extraction form. The following information was extracted: (1) general study characteristics (ie, country of study, study type, participant population, participant demographics, and study sample size); (2) intervention characteristics (ie, intervention name, study design, intervention duration, behavioral theories used, and intervention goals); (3) just-in-time feedback characteristics (eg, content, delivery frequency, and delivery mode); and (4) intervention results (ie, within-group changes and between-group comparisons in targeted behaviors and weight outcomes). A copy of the data extraction form is provided as Multimedia Appendix 1. A comparison of the extracted data across reviewers was conducted (SMS and YL). Differences in the extracted data were resolved by a discussion between the expert reviewers (SMS and YL) to complete the dataset.

Assessing each study for risk of bias was performed using the 2010 version of the Quality Assessment Tool for Quantitative Studies [20]. An overall quality score for each study was assigned based on the ratings for six domains: (1) selection bias, (2) study design, (3) confounders, (4) blinding, (5) data collection methods, and (6) withdrawals and dropouts. The information extraction and quality assessment for each study was performed independently by two reviewers (YL and MCR). Discrepancies between reviewers' ratings were resolved through discussions that led to a consensus (YL, MCR, and SMS).

Analysis

A meta-analysis was not possible owing to substantial heterogeneity in study design, study quality, intervention type, and outcome measures, as well as a lack of studies that explicitly tested the efficacy of using feedback as a BCT. Furthermore, the primary and secondary outcomes of the studies varied widely. We limited our summary of results to primary and secondary study outcomes that were specific to changes in PA, dietary intake, or body weight or body composition. Consistent with the design of a previous review [6], two key criteria were used to determine whether a study explicitly tested feedback: (1) the study had at least one treatment group that received feedback that was not confounded with other manipulations (not matched in the control group) and (2) the study included at least one control group or quasi-control group that received no feedback. Data were synthesized narratively rather than quantitatively.

Results

Literature Search

The literature search yielded 4239 studies, of which 909 were duplicates, leaving 3330 articles to be screened for eligibility. A total of 3083 articles were excluded upon title or abstract screening because they were unrelated to diet or PA, had no reported outcomes, were nonintervention studies, or had ineligible feedback features. Thus, 246 full-text articles were assessed for eligibility. After 215 articles that did not meet the inclusion criteria were excluded, 31 studies with a total of 6623 participants were included in the review (see PRISMA diagram, Figure 1).

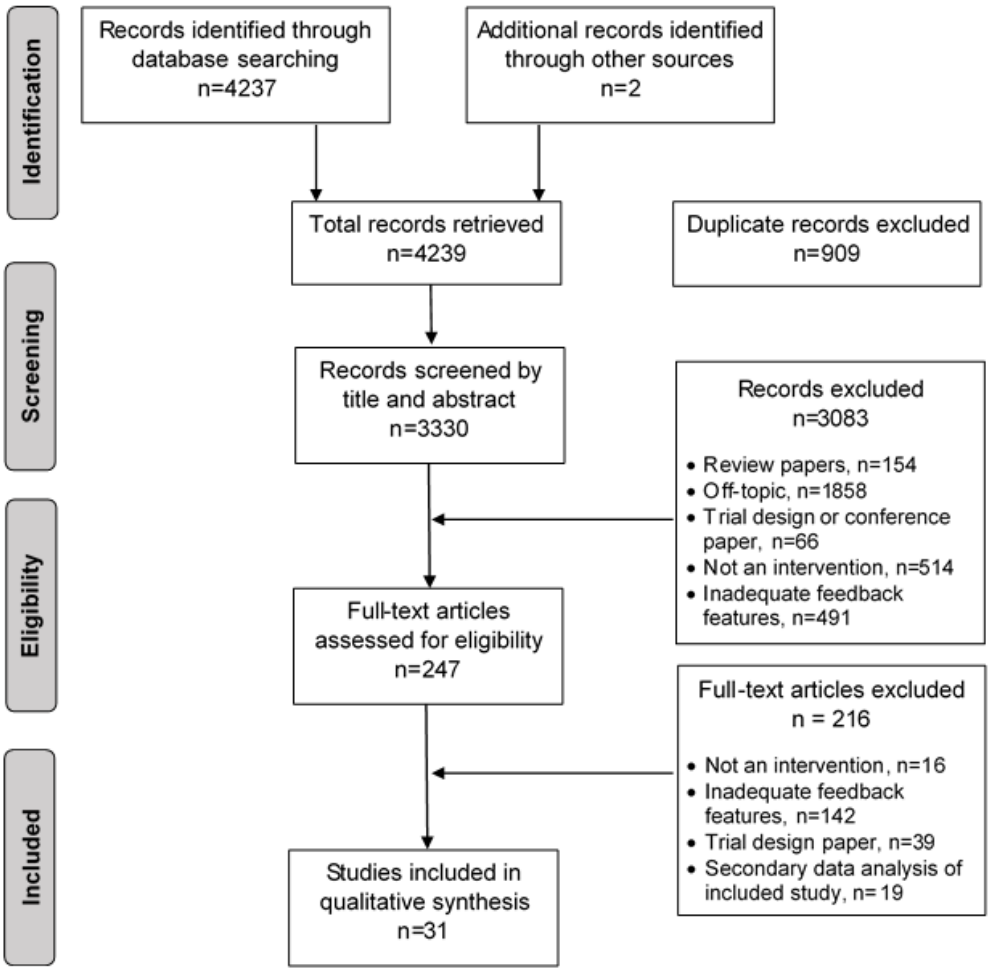
Characteristics of Included Studies

The studies' characteristics are summarized in Multimedia Appendix 2. Studies varied by the behavior about which feedback was provided and by their sample size, population, design, and duration. Of the 31 studies, 3 focused on diet- or nutrition-related behavior only [21-23], 15 focused on PA or sedentary behavior only [24-38], and 13 focused on both diet and PA [39-51]. The median number of study participants was 83 (range=10-1488). Studies were conducted in the United States, the United Kingdom, the Netherlands, Australia, Belgium, Denmark, Hong Kong, Ireland, Japan, Portugal, and South Korea and included adults (≥18 years), children (<18 years), or young adults (17-26 years). The participants' weight statuses were not consistently reported, but at least 11 studies enrolled only overweight or obese individuals [21,22,24,25,39,40,43,47,49-51]. A total of 28 studies were randomized controlled trials (level of evidence I) [21-24,26,27,29-41,43-50], 5 of which used clustered randomization [23,31,32,35,36] (level of evidence I). The 4 remaining studies used within-subjects single-arm [28,42,51] or counterbalanced designs [25] (level of evidence II). Study duration ranged from 2 weeks to 24 months. Four studies [31,32,35,41] had postintervention follow-up periods that ranged from 4 weeks to 6 months.

Risk of Bias

The 31 studies' risks of bias are summarized in Multimedia Appendix 3. Using the current Quality Assessment Tool for Quantitative Studies [20], we determined that 18 studies had a moderate global rating, 9 had a weak global rating, and 4 had a strong global rating. All but 4 studies received a strong study design rating for being randomized controlled trials. A total of 25 studies received strong scores for controlling for potentially confounding variables, 23 studies used data collection measures with demonstrated reliability and validity, and 19 studies had retention rates of ≥80% across conditions. The risk of selection bias posed the greatest threat to validity; 22 studies received a weak score in this domain. Blinding was rated as weak in 4 studies; however, the assessment tool we used may have underestimated this bias [52]. Most studies did not describe blinding procedures for research staff or participants.

Figure 1. Preferred Reporting of Systematic Reviews and Meta-Analyses (PRISMA) diagram.



Feedback Features of Included Studies

The 31 studies varied in the content, frequency, timing, and delivery of feedback, as well as in the methods used to monitor targeted behaviors and the theoretical foundations that guided the feedback content. Multimedia Appendix 4 provides a full description of the feedback features the studies used and a summary of the key features described below.

Theoretical Foundation

In 24 studies [22-24,26,27,29-31,33-44,46,48-50], behavior change theories or guiding principles were used as the foundation for the use of feedback as a BCT. The most frequently endorsed behavior change theories was the SCT (9 studies) [23,27,29,31,39-41,48,49]. Other theories that were endorsed by at least two studies included the control theory [33,34], the health belief model [31], and the transtheoretical model or Stages of Change [38,49]. A total of 8 studies [22,30,31,35-37,42,43] reported using a combination of three or more theories or guiding principles. A total of 7 studies [21,25,28,32,45,47,51] did not specify a guiding theory.

Feedback Content

By its design, this review included only studies that delivered just-in-time feedback. In 10 studies [21,25,30,31, 33,35-37,43,49], feedback was available continuously, in 3 [27,45,46], feedback was provided at multiple times daily, and in 18 [22-24,26,28,29,32,34,38-42,44,47,48,50,51,53], feedback was provided once daily. All but one study [32], which provided team-based feedback, gave *personalized* feedback that reflected the person's own performance. Feedback also often included performance summary information; however, not all feedback was *goal-oriented*. A total of 24 studies [21-25,27-30, 33-35,37-41,44-48,50,51] provided graphical or other visual or verbal feedback on performance relative to known goals or targets, and 8 studies [26,31,32,36,42,43,49] incorporated only raw performance summaries into the feedback. Of the 24 studies that provided goal-oriented feedback, 12 [21-25,27-29, 33,38,45,46] had either self-selected, incremental, or adaptive goals, and 12 [30,34,35,37,39-41,44,47,48,50,51] utilized static goals, targets, or published recommendations (eg, 10,000 steps per day). Goal-oriented feedback that referenced standardized or adaptive thresholds or recommendations was either cumulative (eg, progress toward daily goals) [22-24,28-30, 34,37-41,44,46-48,50,51] or aimed at motivating health-promoting behavior within the day to achieve or maintain a specific behavioral target or threshold throughout that day (eg, achieving daily dietary fat goals) [21,25,27,33,35,45]. In only 5 studies [21,25,30,39,45], the feedback was *actionable*, meaning that it incorporated instructions that explicitly communicated *when*, *where*, and *how* to enact a goal-directed behavior.

Feedback Prompting and Delivery Modes

A total of 17 studies [22-24,26,27,29,34,39-42,44,46-48,50,51] used user-initiated feedback, in which the user initiated feedback delivery by providing his or her self-monitoring data; 11 studies [21,25,28,30-33,35-38] used a passive form of feedback prompting that did not require user interaction to initiate feedback delivery; and 3 studies [43,45,49] employed both methods. Nearly all the studies used an automated form of feedback delivery. All but one of the 17 studies that implemented user-initiated feedback relied on self-reported data. A total of 9 studies [22,23,34,39,42,44,46,47,51] based feedback on measures of behavior that were self-reported via diaries or Web-based self-monitoring tools, 4 studies [24,26,27,29] relied on self-reported measures from activity monitors (eg, pedometers); and 4 studies [40,41,48,50] used a combination self-reported dietary intake and self-reported measures from an activity monitor. All 11 studies [21,25,28,30-33,35-38] that employed a passive form of feedback prompting used objectively assessed data. Nine studies [25,28,30,32,33,35-38] used Internet- or Wi-Fi-connected activity monitors, and 2 studies [21,31] used a mandometer, a Bluetooth-connected scale that measures eating rate, or a heart rate monitor. The remaining 3 studies [43,45,49] used a combination of user-initiated and passive feedback methods. One study [45] used a glucometer.

Efficacy of Feedback Interventions

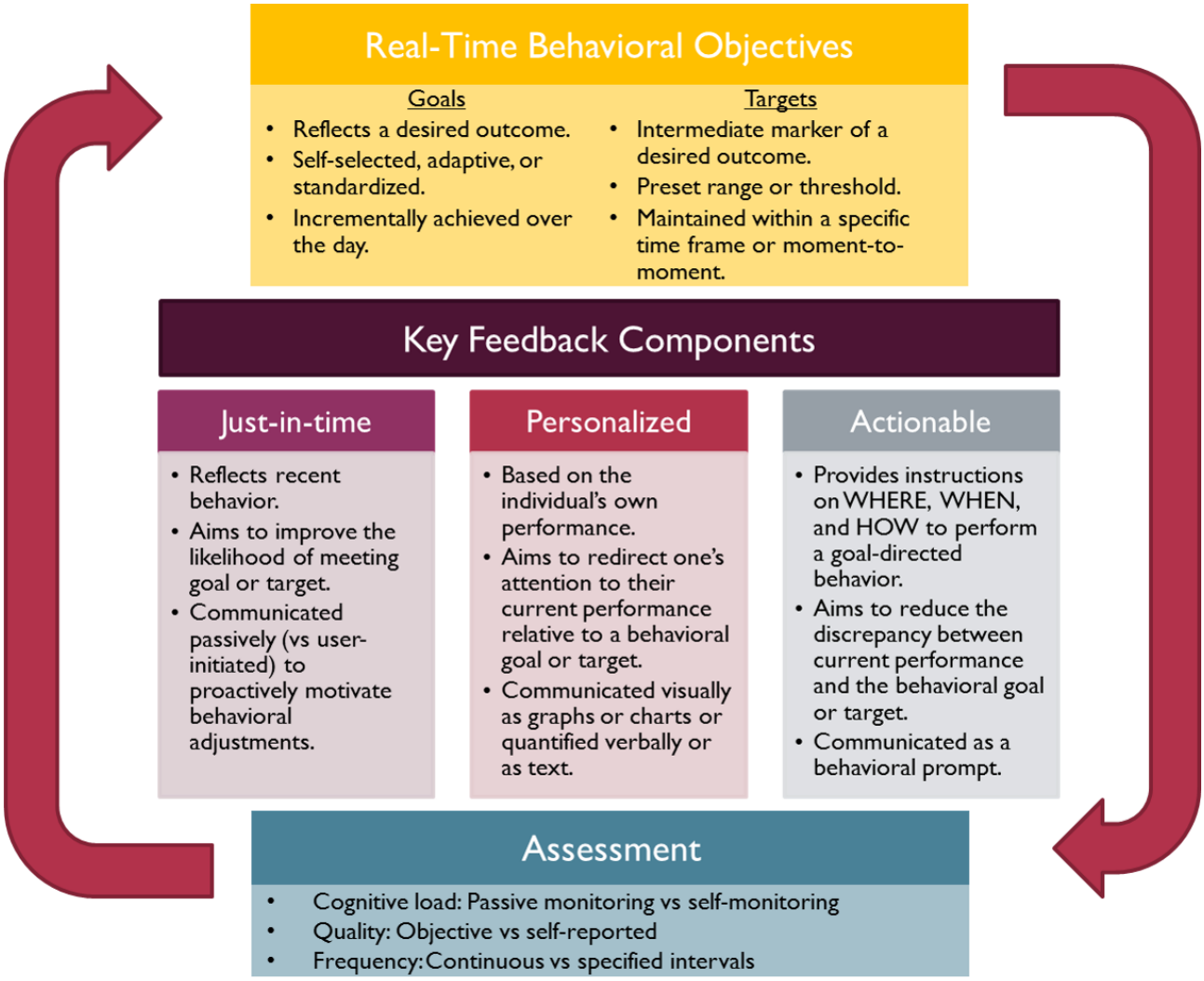
The studies' outcomes are described in Multimedia Appendix 5. We determined that 9 studies explicitly tested the use of feedback to motivate behavior change or significantly modulate body weight or body composition or glycated hemoglobin. Of the 9 studies that tested feedback efficacy, 1 was in the area of diet- or nutrition-related behavior [21], 5 focused on PA or sedentary behavior [26,27,30,34,37], and 3 focused on diet and PA or sedentary behavior [39,45,49]. Most of the remaining studies tested a comprehensive intervention in which feedback was implemented as one of the multiple behavior change strategies. Of the 9 studies that tested the efficacy of feedback, 4 had significant findings [21,30,37,45] and 5 did not [26,27,34,39,49].

Regarding feedback content, 3 of the 4 studies with significant findings [21,30,45] and 3 of the 5 studies without significant findings [27,34,39] used both goal- or target-oriented feedback and actionable feedback. In addition, feedback was provided continuously in 3 of the 4 studies with significant findings [21,30,37] and in only 1 of the 5 studies without significant findings [49]. Concerning feedback prompting, 3 of the 4 studies with significant findings used objectively collected data and passive feedback delivery methods [21,30,37], and the fourth [45] used both passive and user-initiated assessment and feedback methods.

Practical Framework for Just-in-Time Feedback Design

On the basis of the results of our review, we developed a practical framework (Figure 2) that highlights key factors to be considered when developing just-in-time feedback for technology-enhanced diet and PA interventions. We suggest that behavioral objectives (goals or targets) serve as the guiding context for just-in-time feedback and that the selected method of behavioral assessment is the foundation that determines the level of interaction between a user or participant and an external agent or researcher. We propose three characteristics of feedback to take into account: timeliness, personalization, and action orientation.

Figure 2. Practical framework for designing just-in-time feedback.



Behavioral Objectives

Most behavioral theories posit that setting a goal is a key behavior change strategy. Behavioral objectives can be framed as goals or targets. Behavioral goals reflect a desired outcome that is achieved incrementally (eg, over a day); can be static (eg, 10,000 steps per day), or adaptive (adjusted incrementally based on performance achievements) and can be self-selected or assigned by an interventionist or health care professional. Comparatively, behavioral targets can be considered intermediate markers or behavioral mediators of achieving a goal that may be explicitly stated or implicitly understood. For example, a behavioral target could be to take a 5-min activity break after sitting for an hour (ie, reduce sedentary behavior) in an effort to achieve a 10,000-steps-per-day goal (ie, increase PA). Thus, behavioral targets are often set at shorter intervals (ie, within the day) than are goals to keep individuals on track for goal attainment. Behavioral targets can also refer to maintaining a continuously assessed marker within a certain predefined range from moment to moment (ie, blood glucose levels) and can be set at moment-to-moment level to achieve an often implicitly understood distal goal (ie, glycemic control). The chosen behavioral objective determines the context for which the feedback content is designed, as well as the appropriate time frame in which it is delivered.

Behavioral Assessment

The method used to measure the behavior on which feedback is being provided is an important factor for interventions that incorporate just-in-time feedback. The assessment method determines the cognitive load of self-monitoring, the quality of the data, and the frequency and timeliness with which feedback is delivered. Advancements in wearable sensor technology influence these aspects of data collection, particularly in the area of PA behavior. Newer activity monitors can not only objectively and continuously measure movement (eg, steps) and estimate energy expenditures but also transfer data to another device (eg, mobile phones) or servers wirelessly. This enables the generation and delivery of feedback without any user-initiated input.

Timeliness

Just-in-time feedback is defined as providing the right support at the right moment and in the right amount [5,11,12]. In just-in-time adaptive interventions, the right moment might consider the person's state of vulnerability or opportunity or receptivity. Consistent with the FIT, just-in-time feedback reflects recent behavior and provides guidance at a critical

moment when a person's attention might need to be refocused on the goal-directed behavior. The intent of delivering just-in-time feedback is to proactively motivate behavioral adjustments to limit or reverse the widening of discrepancies between current performance and a behavioral goal or target to facilitate the attainment of that goal or target. As such, just-in-time feedback considers the time frame during which goal achievement is specified to occur. For example, if one has a goal to achieve 10,000 steps per day, just-in-time feedback would be provided before the end of the day to increase the likelihood of achieving that day's step goal.

Personalization

Rather than being generic or group-based, personalized feedback is based on an individual's own performance and goal. The intent of personalizing feedback is to inform a person about his or her current performance relative to his or her behavioral goal or target (ie, the discrepancy). This message can be communicated visually as graphs or charts (eg, a progress bar) or quantified as text (eg, 3000 more steps to meet your goal) to meet the research needs or the user's preference.

Action Orientation

Feedback that is actionable aims to instruct a patient or participant to engage in behaviors that will improve the likelihood of goal attainment. Action plans are designed to promote small or large behavior changes with a high likelihood of success and should indicate *when*, *where*, and *how* to enact a goal-directed behavior [54]. Actionable feedback provides behavioral guidance aimed at reducing the discrepancy between current performance and a behavioral goal or target. In just-in-time interventions, action planning varies from its traditional sense of intention formation [54] such that it can be communicated explicitly as instructions for behaviors to be enacted at critical moments [55,56]. These instructions act as behavioral triggers and can be communicated through prompts (eg, Try going for a 30-min walk after dinner tonight to get the 2500 more steps you need to meet your 10,000 step goal today).

Discussion

Summary of Key Findings

Our comprehensive literature search yielded 31 studies that met our eligibility criteria. Most of these studies provided feedback that was *just-in-time* (100%, 31/31), *goal-oriented* (75%, 23/31), and *personalized* (97%, 30/31); however, only 5 studies (15%, 5/31) provided *actionable* feedback. Interventions with significant effects featured continuously available and personalized feedback that was actionable or goal-oriented and used objectively assessed data and passively initiated feedback delivery methods. On the basis of these findings, we have proposed a framework for designing just-in-time feedback that incorporates three key content characteristics (timeliness, personalization, and action orientation) relative to a known behavioral goal or target and highlights important considerations regarding the quality and frequency of the behavioral assessment from which just-in-time feedback is derived.

Comparison With Previous Literature

To our knowledge, our review is unique in that its primary goal was to examine the use of just-in-time feedback in diet and PA interventions. Only one other similar systematic review has been conducted. In that review [17], the authors concluded that the use of personalized feedback was an effective BCT in weight-loss interventions after demonstrating that individuals receiving Internet-delivered personalized feedback lost on average 2 kg more than those receiving no personalized feedback; however, the authors did not analyze the effectiveness of other feedback characteristics. In another review, Michie and colleagues [56] demonstrated through a meta-regression analysis that healthy eating and PA interventions that implemented behavioral monitoring plus one BCT (eg, action planning, prompting specific goal setting, providing feedback on performance, and prompting review of behavioral goals) were more effective than those that did not. However, the authors acknowledged the need to experimentally test the most effective combination of BCTs [56]. Furthermore, only one other known model for developing feedback exists. Hysong and colleagues [57] examined the use of feedback for improving clinical practice guideline adherence and developed the model of actionable feedback. The model posits that an optimal effect on clinical performance can be achieved by providing feedback that is timely, individualized, nonpunitive, and based on customizable performance data. Action was an implied outcome of providing optimal feedback in the model. These previous studies highlight some potentially effective components of feedback that can motivate goal-directed behavior change and that overlap with those put forth in this review.

Another key finding of our review was that only a few of the reviewed studies provided goal-oriented feedback that was actionable. Of the 31 studies included in the review, 23 incorporated behavioral goals; however, only 5 [21,25,30,39,45] gave actionable suggestions on when, where, and how to achieve those goals. In theory, intention precedes action [58]; however, research shows that the intention alone does not often result in actual behavior change [59-61]. Action planning or intention implementation is one strategy to help people transform their intention into action [54]. Traditionally, individuals conceive action plans before acting as their commitment to perform behaviors when opportunities arise [54] and can reevaluate these plans weekly [62]. Such plans often describe where, when, or how frequently to perform the behavior (eg, I intend go to the gym 3 days per week) [55]. Individuals may or may not receive feedback regarding their action plans. In this study *actionable* or *actionable feedback* refers more specifically to information in the form of suggestions an external agent provides to an individual about *where* and *when* to perform the behavior and instructions on *how* to perform the behavior. Examples of actionable feedback in the reviewed studies included multiple, daily behavioral prompts to perform PA breaks of specific durations [25], daily booster messages with exercise prescriptions to achieve the current day's step goal [30], and just-in-time dietary recommendations (eg, Be aware of high fat snacks tonight) [39]. Our definition of actionable feedback is consistent with Michie and colleagues' Coventry, Aberdeen, and London-Refined (CALO-RE) taxonomy of BCTs [55]. It

is also consistent with the concept of providing supportive information, advice, and feedback at critical moments in just-in-time adaptive interventions [5,11]. This review demonstrates that the provision of actionable feedback in diet and PA interventions is an underutilized behavior change strategy. Relatedly, reviews of seven wearable activity trackers and 40 top-rated diet and PA mobile phone apps found that the integration of BCTs related to action planning and providing instructions consistent with CALO-RE taxonomy was not uncommon [63,64]. However, only two activity trackers helped users identify *where* and *when* to perform PA. Given the increased bidirectional communication capabilities offered by the technology being used to facilitate behavioral interventions, actionable feedback as defined here will likely be more frequently incorporated into interventions as a supportive behavior change strategy. One example of such an intervention is the MyBehavior mobile app. MyBehavior is based on the Fogg behavioral model that applies theoretical principles to technology design by creating tools to prompt low-effort actions that can be triggered even when motivation is low [65]. MyBehavior was designed to generate personalized, actionable insights on when, where, and how to achieve the set goals [66]. The behavioral outcomes from the MyBehavior trial have not yet been published. Future research to determine the efficacy of actionable versus nonactionable feedback as a behavior change strategy will be needed.

Strengths and Limitations

Our review is strengthened by its focus on key theory-based characteristics of feedback delivered as behavior change interventions. We focused on only diet and PA interventions rather than looking more broadly across additional health behaviors. We did this in part because it is unclear how generalizable our findings might be to other behaviors or health-related outcomes. In addition, studies eligible for inclusion could have included multiple types of feedback, but at least one form of feedback had to meet our definition of just-in-time feedback. We believe this approach strengthens this review by enabling us to make conclusions that facilitate the progress of intervention science into a future in which feedback can be generated and delivered just in time, thereby preparing researchers for continued advancements in technology. Finally, our synthesis of the available data enabled us to develop a framework for designing just-in-time feedback for health behavior change interventions.

Despite these strengths, we were unable to conduct a meta-analysis primarily because of the variability in targeted behavior and study outcomes. Additionally, because it was not clear whether the included studies monitored the delivery, receipt or viewing, or comprehension of the provided feedback, we were not able to conclusively determine the efficacy of using feedback or which feedback feature(s) might be more effective than others. However, we found that feedback was continuously available, goal-oriented, or actionable in 3 of the 4 studies with significant intervention effects. In addition, the sample sizes, intervention durations, and interventions outcomes of studies with significant findings ranged widely. Interventions of longer duration could have been more likely to have significant findings. However, the duration of studies with significant findings was generally shorter than studies with nonsignificant findings (4 weeks to 12 months vs 2 weeks to 24 months). Excluding studies for providing feedback more than 24 hours after a person performed the target behavior limited the number of eligible studies; however, we believe that it was consistent with the advancement of body sensor technology and therefore important to the context of the review. Another important limitation to consider is that most of the included studies did not recruit participants from a representative, diverse population, thus limiting the generalizability of the findings. Finally, the cost of wearable body sensors and wireless devices could be a potential limitation for scaling up the technology-based interventions.

Implications for Future Research

Before this study, few reviews had critically examined the use of feedback in diet and PA interventions. As advancements in technology continue to improve bidirectional communication between investigators and their participants, optimizing feedback messages will be key to future interventions. The systematic review and the framework we propose represent a foundation for designing feedback messages for future just-in-time diet- or PA-based interventions. Investigators may use the framework to ensure feedback developed for their interventions contain content that is theoretically and empirically supported to have a positive effect on behavior change. However, it is unclear from this review how many of the proposed components are needed to effectively motivate behavior change. Empirical research will be needed to determine the optimal combination of feedback components.

Acknowledgments

The authors acknowledge Catalina Malinowski for her assistance with article review and data extraction. The preparation of this manuscript was supported in part by NIH grants R21CA215415, R01HL119255, R01CA186700, R25ECA056452, and P30CA016672; a Janice Davis Gordon Memorial Postdoctoral Fellowship in Colorectal Cancer Prevention; an Institutional Research Grant from The University of Texas MD Anderson Cancer Center; the Chandler Cox Foundation; the Duncan Family Institute for Cancer Prevention and Risk Assessment; and the Center for Energy Balance in Cancer Prevention and Survivorship. The supporting sources had no involvement in the study design; collection, analysis, and interpretation of data; writing the report; or the decision to submit the report for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Data extraction form.

[PDF File (Adobe PDF File), 21KB - jmir_v20i3e106_app1.pdf]

Multimedia Appendix 2

Characteristics of included intervention studies.

[PDF File (Adobe PDF File), 58KB - jmir_v20i3e106_app2.pdf]

Multimedia Appendix 3

Risk of bias summaries for included intervention studies.

[PDF File (Adobe PDF File), 51KB - jmir_v20i3e106_app3.pdf]

Multimedia Appendix 4

Descriptions and summaries of feedback features in included intervention studies.

[PDF File (Adobe PDF File), 75KB - jmir_v20i3e106_app4.pdf]

Multimedia Appendix 5

Summaries of outcomes and feedback efficacy for included intervention studies.

[PDF File (Adobe PDF File), 97KB - jmir_v20i3e106_app5.pdf]

References

1. Fox S, Duggan M. Pew Internet. Washington, DC: Pew Research Center's Internet & American Life Project; 2013 Jan 28. Tracking for health URL: http://www.pewinternet.org/2013/01/28/tracking-for-health/ [accessed 2017-05-01] [WebCite Cache ID 6upaSFP3u]
2. Gualtieri L, Rosenbluth S, Phillips J. Can a free wearable activity tracker change behavior? The impact of trackers on adults in a physician-led wellness group. JMIR Res Protoc 2016;5(4):e237 [FREE Full text] [doi: 10.2196/resprot.6534] [Medline: 27903490]
3. Pourzanjani A, Quisel T, Foschini L. Adherent use of digital health trackers is associated with weight loss. PLoS One 2016;11(4):e0152504 [FREE Full text] [doi: 10.1371/journal.pone.0152504] [Medline: 27049859]
4. Pellegrini CA, Verba SD, Otto AD, Helsel DL, Davis KK, Jakicic JM. The comparison of a technology-based system and an in-person behavioral weight loss intervention. Obesity (Silver Spring) 2012 Feb;20(2):356-363. [doi: 10.1038/oby.2011.13] [Medline: 21311506]
5. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (jitais) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2016 Sep 23:1-17 (forthcoming)(forthcoming). [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
6. Kluger AN, DeNisi A. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. Psychol Bull 1996;119(2):254-284. [doi: 10.1037/0033-2909.119.2.254]
7. Carver CS, Scheier MF. Attention and self-regulation: a Control Theory approach to human behavior. New York: Springer; 1982.
8. Carver CS, Scheier MF. Control theory: a useful conceptual framework for personality-social, clinical, and health psychology. Psychol Bull 1982 Jul;92(1):111-135. [Medline: 7134324]
9. Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. Am Psychol 2002 Sep;57(9):705-717. [Medline: 12237980]
10. Bandura A. The evolution of social cognitive theory. In: Smith KG, Hitt MA, editors. Great minds in management: The process of theory development. New York: Oxford University Press; 2005:9-35.
11. Nahum-Shani I, Hekler EB, Spruijt-Metz D. Building health behavior models to guide the development of just-in-time adaptive interventions: a pragmatic framework. Health Psychol 2015 Dec;34 Suppl:1209-1219. [doi: 10.1037/hea0000306] [Medline: 26651462]
12. Spruijt-Metz D, Wen CK, O'Reilly G, Li M, Lee S, Emken BA, et al. Innovations in the use of interactive technology to support weight management. Curr Obes Rep 2015 Dec;4(4):510-519 [FREE Full text] [doi: 10.1007/s13679-015-0183-6] [Medline: 26364308]

13. Gardner B, Whittington C, McAteer J, Eccles MP, Michie S. Using theory to synthesise evidence from behaviour change interventions: the example of audit and feedback. Soc Sci Med 2010 May;70(10):1618-1625. [doi: 10.1016/j.socscimed.2010.01.039] [Medline: 20207464]
14. Bardus M, Smith JR, Samaha L, Abraham C. Mobile phone and web 2.0 technologies for weight management: a systematic scoping review. J Med Internet Res 2015 Nov 16;17(11):e259 [FREE Full text] [doi: 10.2196/jmir.5129] [Medline: 26573984]
15. Müller AM, Alley S, Schoeppe S, Vandelanotte C. The effectiveness of e-& mHealth interventions to promote physical activity and healthy diets in developing countries: a systematic review. Int J Behav Nutr Phys Act 2016 Oct 10;13(1):109 [FREE Full text] [doi: 10.1186/s12966-016-0434-2] [Medline: 27724911]
16. Thomas JG, Bond DS. Review of innovations in digital health technology to promote weight control. Curr Diab Rep 2014;14(5):485. [doi: 10.1007/s11892-014-0485-1] [Medline: 24664797]
17. Sherrington A, Newham JJ, Bell R, Adamson A, McColl E, Araujo-Soares V. Systematic review and meta-analysis of internet-delivered interventions providing personalized feedback for weight loss in overweight and obese adults. Obes Rev 2016 Jun;17(6):541-551 [FREE Full text] [doi: 10.1111/obr.12396] [Medline: 26948257]
18. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol 2009 Oct;62(10):e1-34 [FREE Full text] [doi: 10.1016/j.jclinepi.2009.06.006] [Medline: 19631507]
19. Higgins J, Green S. Cochrane handbook for systematic reviews of interventions: Cochrane book series. Chichester, UK: Wiley-Blackwell; Sep 2008.
20. Thomas BH, Ciliska D, Dobbins M, Micucci S. A process for systematically reviewing the literature: providing the research evidence for public health nursing interventions. Worldviews Evid Based Nurs 2004;1(3):176-184. [doi: 10.1111/j.1524-475X.2004.04006.x] [Medline: 17163895]
21. Ford AL, Bergh C, Södersten P, Sabin MA, Hollinghurst S, Hunt LP, et al. Treatment of childhood obesity by retraining eating behaviour: randomised controlled trial. Br Med J 2009 Jan 05;340:b5388. [doi: 10.1136/bmj.b5388] [Medline: 20051465]
22. Mummah SA, Mathur M, King AC, Gardner CD, Sutton S. Mobile technology for vegetable consumption: a randomized controlled pilot study in overweight adults. JMIR Mhealth Uhealth 2016 May;4(2):e51 [FREE Full text] [doi: 10.2196/mhealth.5146] [Medline: 27193036]
23. Pedersen S, Grønhøj A, Thøgersen J. Texting your way to healthier eating? Effects of participating in a feedback intervention using text messaging on adolescents' fruit and vegetable intake. Health Educ Res 2016 Apr;31(2):171-184. [doi: 10.1093/her/cyv104] [Medline: 26850061]
24. 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;8(12):e82901 [FREE Full text] [doi: 10.1371/journal.pone.0082901] [Medline: 24349392]
25. Bond DS, Thomas JG, Raynor HA, Moon J, Sieling J, Trautvetter J, et al. B-MOBILE--a smartphone-based intervention to reduce sedentary time in overweight/obese individuals: a within-subjects experimental trial. PLoS One 2014;9(6):e100821 [FREE Full text] [doi: 10.1371/journal.pone.0100821] [Medline: 24964010]
26. Chapman GB, Colby H, Convery K, Coups EJ. Goals and social comparisons promote walking behavior. Med Decis Making 2016 May;36(4):472-478. [doi: 10.1177/0272989X15592156] [Medline: 26139447]
27. 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]
28. Hooke MC, Gilchrist L, Tanner L, Hart N, Withycombe JS. Use of a fitness tracker to promote physical activity in children with acute lymphoblastic leukemia. Pediatr Blood Cancer 2016 Apr;63(4):684-689. [doi: 10.1002/pbc.25860] [Medline: 26756736]
29. King AC, Ahn DK, Oliveira BM, Atienza AA, Castro CM, Gardner CD. Promoting physical activity through hand-held computer technology. Am J Prev Med 2008 Feb;34(2):138-142 [FREE Full text] [doi: 10.1016/j.amepre.2007.09.025] [Medline: 18201644]
30. Martin SS, Feldman DI, Blumenthal RS, Jones SR, Post WS, McKibben RA, et al. Mactive: a randomized clinical trial of an automated mHealth intervention for physical activity promotion. J Am Heart Assoc 2015 Nov 09;4(11):1-9 [FREE Full text] [doi: 10.1161/JAHA.115.002239] [Medline: 26553211]
31. McManus AM, Masters RS, Laukkanen RM, Yu CC, Sit CH, Ling FC. Using heart-rate feedback to increase physical activity in children. Prev Med 2008 Oct;47(4):402-408. [doi: 10.1016/j.ypmed.2008.06.001] [Medline: 18590757]
32. Patel MS, Volpp KG, Rosin R, Bellamy SL, Small DS, Fletcher MA, et al. A randomized trial of social comparison feedback and financial incentives to increase physical activity. Am J Health Promot 2016 Jul;30(6):416-424. [doi: 10.1177/0890117116658195] [Medline: 27422252]
33. Paul L, Wyke S, Brewster S, Sattar N, Gill JM, Alexander G, et al. Increasing physical activity in stroke survivors using STARFISH, an interactive mobile phone application: a pilot study. Top Stroke Rehabil 2016 Jun;23(3):170-177. [doi: 10.1080/10749357.2015.1122266] [Medline: 27077973]

34. Prestwich A, Conner M, Hurling R, Ayres K, Morris B. An experimental test of control theory-based interventions for physical activity. Br J Health Psychol 2016 Nov;21(4):812-826. [doi: 10.1111/bjhp.12198] [Medline: 27169809]
35. van der Weegen S, Verwey R, Spreeuwenberg M, Tange H, van der Weijden T, de Witte L. It's life! Mobile and web-based monitoring and feedback tool embedded in primary care increases physical activity: a cluster randomized controlled trial. J Med Internet Res 2015 Jul 24;17(7):e184 [FREE Full text] [doi: 10.2196/jmir.4579] [Medline: 26209025]
36. van Nimwegen M, Speelman AD, Overeem S, van de Warrenburg BP, Smulders K, Dontje ML, et al. Promotion of physical activity and fitness in sedentary patients with Parkinson's disease: randomised controlled trial. Br Med J 2013 Mar 01;346:f576 [FREE Full text] [Medline: 23457213]
37. Walsh JC, Corbett T, Hogan M, Duggan J, McNamara A. An mhealth intervention using a smartphone app to increase walking behavior in young adults: a pilot study. JMIR Mhealth Uhealth 2016 Sep 22;4(3):e109 [FREE Full text] [doi: 10.2196/mhealth.5227] [Medline: 27658677]
38. Wijsman CA, Westendorp RG, Verhagen EA, Catt M, Slagboom PE, de Craen AJ, et al. Effects of a web-based intervention on physical activity and metabolism in older adults: randomized controlled trial. J Med Internet Res 2013;15(11):e233 [FREE Full text] [doi: 10.2196/jmir.2843] [Medline: 24195965]
39. Burke LE, Styn MA, Sereika SM, Conroy MB, Ye L, Glanz K, et al. Using mHealth technology to enhance self-monitoring for weight loss: a randomized trial. Am J Prev Med 2012 Jul;43(1):20-26 [FREE Full text] [doi: 10.1016/j.amepre.2012.03.016] [Medline: 22704741]
40. Collins CE, Morgan PJ, Jones P, Fletcher K, Martin J, Aguiar EJ, et al. A 12-week commercial web-based weight-loss program for overweight and obese adults: randomized controlled trial comparing basic versus enhanced features. J Med Internet Res 2012;14(2):e57 [FREE Full text] [doi: 10.2196/jmir.1980] [Medline: 22555246]
41. Fassnacht DB, Ali K, Silva C, Gonçalves S, Machado PP. Use of text messaging services to promote health behaviors in children. J Nutr Educ Behav 2015;47(1):75-80. [doi: 10.1016/j.jneb.2014.08.006] [Medline: 25282200]
42. Fujii H, Nakade M, Haruyama Y, Fukuda H, Hashimoto M, Ikuyama T, et al. Evaluation of a computer-tailored lifestyle modification support tool for employees in Japan. Ind Health 2009 Jul;47(3):333-341 [FREE Full text] [Medline: 19531920]
43. 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. J Am Med Assoc 2016 Sep 20;316(11):1161-1171. [doi: 10.1001/jama.2016.12858] [Medline: 27654602]
44. Khanna R, Stoddard PJ, Gonzales EN, Villagran-Flores M, Thomson J, Bayard P, et al. An automated telephone nutrition support system for Spanish-speaking patients with diabetes. J Diabetes Sci Technol 2014 Nov;8(6):1115-1120 [FREE Full text] [doi: 10.1177/1932296814550186] [Medline: 25239122]
45. Lim S, Kang SM, Kim KM, Moon JH, Choi SH, Hwang H, et al. Multifactorial intervention in diabetes care using real-time monitoring and tailored feedback in type 2 diabetes. Acta Diabetol 2015 Apr;53(2):189-198. [doi: 10.1007/s00592-015-0754-8] [Medline: 25936739]
46. Nollen NL, Mayo MS, Carlson SE, Rapoff MA, Goggin KJ, Ellerbeck EF. Mobile technology for obesity prevention: a randomized pilot study in racial- and ethnic-minority girls. Am J Prev Med 2014 Apr;46(4):404-408 [FREE Full text] [doi: 10.1016/j.amepre.2013.12.011] [Medline: 24650843]
47. Sacks FM, Bray GA, Carey VJ, Smith SR, Ryan DH, Anton SD, et al. Comparison of weight-loss diets with different compositions of fat, protein, and carbohydrates. N Engl J Med 2009 Feb 26;360(9):859-873 [FREE Full text] [doi: 10.1056/NEJMoa0804748] [Medline: 19246357]
48. Shapiro JR, Bauer S, Hamer RM, Kordy H, Ward D, Bulik CM. Use of text messaging for monitoring sugar-sweetened beverages, physical activity, and screen time in children: a pilot study. J Nutr Educ Behav 2008;40(6):385-391 [FREE Full text] [doi: 10.1016/j.jneb.2007.09.014] [Medline: 18984496]
49. Shuger SL, Barry VW, Sui X, McClain A, Hand GA, Wilcox S, et al. Electronic feedback in a diet- and physical activity-based lifestyle intervention for weight loss: a randomized controlled trial. Int J Behav Nutr Phys Act 2011 May 18;8:41 [FREE Full text] [doi: 10.1186/1479-5868-8-41] [Medline: 21592351]
50. 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;15(11):e244 [FREE Full text] [doi: 10.2196/jmir.2844] [Medline: 24246427]
51. Willey S, Walsh JK. Outcomes of a mobile health coaching platform: 12-week results of a single-arm longitudinal study. JMIR Mhealth Uhealth 2016 Jan;4(1):e3 [FREE Full text] [doi: 10.2196/mhealth.4933] [Medline: 26747611]
52. Armijo-Olivo S, Stiles CR, Hagen NA, Biondo PD, Cummings GG. Assessment of study quality for systematic reviews: a comparison of the cochrane collaboration risk of bias tool and the effective public health practice project quality assessment tool: methodological research. J Eval Clin Pract 2012 Feb;18(1):12-18. [doi: 10.1111/j.1365-2753.2010.01516.x] [Medline: 20698919]
53. Richardson CR, Mehari KS, McIntyre LG, Janney AW, Fortlage LA, Sen A, et al. A randomized trial comparing structured and lifestyle goals in an internet-mediated walking program for people with type 2 diabetes. Int J Behav Nutr Phys Act 2007 Nov 16;4:59 [FREE Full text] [doi: 10.1186/1479-5868-4-59] [Medline: 18021411]
54. Gollwitzer PM. Implementation intentions: strong effects of simple plans. Am Psychol 1999 Jul;54(7):493-503. [doi: 10.1037//0003-066x.54.7.493]

55. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
56. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
57. Hysong SJ, Best RG, Pugh JA. Audit and feedback and clinical practice guideline adherence: making feedback actionable. Implement Sci 2006 Apr 28;1:9 [FREE Full text] [doi: 10.1186/1748-5908-1-9] [Medline: 16722539]
58. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
59. Rhodes RE, Dickau L. Experimental evidence for the intention-behavior relationship in the physical activity domain: a meta-analysis. Health Psychol 2012 Nov;31(6):724-727. [doi: 10.1037/a0027290] [Medline: 22390739]
60. Hagger M, Chatzisarantis N, Biddle S. A meta-analytic review of the theories of reasoned action and planned behavior in physical activity: predictive validity and the contribution of additional variables. J Sport Exerc Psychol 2002 Mar;24(1):3-32. [doi: 10.1123/jsep.24.1.3]
61. Hagger MS, Luszczynska A. Implementation intention and action planning interventions in health contexts: state of the research and proposals for the way forward. Appl Psychol Health Well Being 2014 Mar;6(1):1-47. [doi: 10.1111/aphw.12017] [Medline: 24591064]
62. Lorig KR, Sobel DS, Stewart AL, Brown BW, Bandura A, Ritter P, et al. Evidence suggesting that a chronic disease self-management program can improve health status while reducing hospitalization: a randomized trial. Med Care 1999 Jan;37(1):5-14. [Medline: 10413387]
63. Mercer K, Li M, Giangregorio L, Burns C, Grindrod K. Behavior change techniques present in wearable activity trackers: a critical analysis. JMIR Mhealth Uhealth 2016 Apr 27;4(2):e40 [FREE Full text] [doi: 10.2196/mhealth.4461] [Medline: 27122452]
64. Direito A, Dale LP, Shields E, Dobson R, Whittaker R, Maddison R. Do physical activity and dietary smartphone applications incorporate evidence-based behaviour change techniques? BMC Public Health 2014 Jun 25;14(646):1-7 [FREE Full text] [doi: 10.1186/1471-2458-14-646] [Medline: 24965805]
65. Fogg B. A behavior model for persuasive design. In: Proceedings of the 4th International Conference on Persuasive Technology. New York, NY: Association for Computing Machinery; 2009 Apr Presented at: Persuasive '09; April 26-29, 2009; Claremont, CA. [doi: 10.1145/1541948.1541999]
66. 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;3(2):e42 [FREE Full text] [doi: 10.2196/mhealth.4160] [Medline: 25977197]

Abbreviations

BCT: behavior change technique
FIT: feedback intervention theory
IG: intervention group
PA: physical activity
PDA: personal digital assistant
PRISMA: Preferred Reporting of Systematic Reviews and Meta-Analyses
SCT: social cognitive theory

Edited by G Eysenbach; submitted 10.08.17; peer-reviewed by B Turner-McGrievy, IN Gomez; comments to author 18.09.17; revised version received 19.12.17; accepted 22.12.17; published 22.03.18.

Please cite as:
Schembre SM, Liao Y, Robertson MC, Dunton GF, Kerr J, Haffey ME, Burnett T, Basen-Engquist K, Hicklen RS
Just-in-Time Feedback in Diet and Physical Activity Interventions: Systematic Review and Practical Design Framework
J Med Internet Res 2018;20(3):e106
URL: http://www.jmir.org/2018/3/e106/
doi:10.2196/jmir.8701
PMID:29567638

Original Paper

Transition of a Text-Based Insulin Titration Program From a Randomized Controlled Trial Into Real-World Settings: Implementation Study

Natalie Koch Levy[1], MD; Natasha A Orzeck-Byrnes[1], BA; Sneha R Aidasani[1], MS; Dana N Moloney[1], MA; Lisa H Nguyen[1], MS; Agnes Park[1], BS; Lu Hu[2], PhD; Aisha T Langford[2], PhD; Binhuan Wang[2], PhD; Mary Ann Sevick[2], ScD; Erin S Rogers[2], Dr PH

[1]Division of General Internal Medicine and Clinical Innovation, Department of Medicine, New York University School of Medicine, New York, NY, United States

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

Corresponding Author:
Natalie Koch Levy, MD
Division of General Internal Medicine and Clinical Innovation
Department of Medicine
New York University School of Medicine
OBVA 616
462 First Avenue
New York, NY, 10016
United States
Phone: 1 212 263 8924
Fax: 1 212 263 8788
Email: natalie.levy@nyumc.org

Abstract

Background: The Mobile Insulin Titration Intervention (MITI) program helps patients with type 2 diabetes find their correct basal insulin dose without in-person care. Requiring only basic cell phone technology (text messages and phone calls), MITI is highly accessible to patients receiving care in safety-net settings. MITI was shown in a randomized controlled trial (RCT) to be efficacious at a New York City (NYC) safety-net clinic where patients often have challenges coming for in-person care. In 2016, MITI was implemented as usual care at Bellevue Hospital (the site of the original RCT) and at Gouverneur Health (a second NYC safety-net clinic) under 2 different staffing models.

Objective: This implementation study examined MITI's transition into real-world settings. To understand MITI's flexibility, generalizability, and acceptability among patients and providers, we evaluated whether MITI continued to produce positive outcomes in expanded underserved populations, outside of an RCT setting.

Methods: Patients enrolled in MITI received weekday text messages asking for their fasting blood glucose (FBG) values and a weekly titration call. The goal was for patients to reach their optimal insulin dose (OID), defined either as the dose of once-daily basal insulin required to achieve either an FBG of 80-130 mg/dL (4.4-7.2 mmol/L) or as the reaching of the maximum dose of 50 units. After 12 weeks, if OID was not reached, the patients were asked to return to the clinic for in-person care and titration. MITI program outcomes, clinical outcomes, process outcomes, and patient satisfaction were assessed.

Results: MITI was successful at both sites, each with a different staffing model. Providers referred 170 patients to the program—129 of whom (75.9%, 129/170) were eligible. Of these, 113 (87.6%, 113/129) enrolled. Moreover, 84.1% (95/113) of patients reached their OID, and they did so in an average of 24 days. Clinical outcomes show that mean FBG levels fell from 209 mg/dL (11.6 mmol/L) to 141 mg/dL (7.8 mmol/L), $P<.001$. HbA_{1c} levels fell from 11.4% (101 mmol/mol) to 10.0% (86 mmol/mol), $P<.001$. Process outcomes show that 90.1% of MITI's text message prompts received a response, nurses connected with patients 81.9% of weeks to provide titration instructions, and 85% of attending physicians made at least one referral to the MITI program. Satisfaction surveys showed that most patients felt comfortable sharing information over text and felt the texts reminded them to take their insulin, check their sugar, and make healthy food choices.

Conclusions: This implementation study showed MITI to have continued success after transitioning from an RCT program into real-world settings. MITI showed itself to be flexible and generalizable as it easily fits into a second site staffed by general medical clinic–registered nurses and remained acceptable to patients and staff who had high levels of engagement with the program.

(J Med Internet Res 2018;20(3):e93) doi:10.2196/jmir.9515

KEYWORDS

insulin/long-acting/administration & dosage; diabetes mellitus, type 2/drug therapy; medically underserved area; telemedicine; healthcare disparities

Introduction

Background

According to the Centers for Disease Control and Prevention, as of 2015, 30.3 million people in the United States (9.4% of the population) had diabetes [1]. By 2050, the prevalence of the disease is expected to rise, affecting as many as 1 in 3 Americans [2]. Moreover, diabetes is one of the costliest diseases in the United States [3]. In 2012, the total cost associated with the disease was US $245 billion, including US $176 billion in direct medical care costs [4]. Diabetes puts patients at risk for blindness, kidney failure, nontraumatic limb amputation, stroke, and coronary heart disease [5-10]. Controlling blood sugar can help prevent the long-term adverse health consequences of diabetes [7-11].

Prior research has found significant disparities in the burden of diabetes, such that the prevalence is much higher in racial and ethnic minorities than whites [1]. Compounding the disparity in disease burden, racial and ethnic minorities have less access to care than whites [12]. Taken together, racial and ethnic minorities suffer disproportionately from the negative health consequences of uncontrolled diabetes [1,12,13].

The majority of patients with diabetes can control their blood sugar with lifestyle modifications and oral medications. However, more than 25% of patients in the United States with diabetes take insulin [14]. When insulin is needed, the first type of insulin started is basal insulin, which is in charge of controlling the blood sugar in the fasting state, for example, while we sleep. Patients with type 2 diabetes starting on a given dose of basal insulin are asked to check their morning fasting blood sugar at home daily while taking that particular dose [15]. Then, the dose of basal insulin is gradually increased (or *titrated*) until the morning fasting blood sugar falls into the desired range [15]. Ideally, a patient's dose of basal insulin will bring all of the fasting sugars into the desired range. In reality, patients have variability in factors such as food intake and energy expenditure from day to day, which creates variable fasting blood sugars. It is important to note that basal insulin titration is limited by the lowest sugar. Thus, once any morning fasting blood sugar falls into the desired range, caution must be exercised with any further increase in basal insulin to avoid hypoglycemia.

Of note, in this scenario variability, more insulin may be needed on many days, but it will need to come in the form of a second type of insulin referred to as mealtime insulin (or "bolus") insulin. Unlike the basal insulin discussed above that is given at a consistent dose each night, mealtime insulin is rapid acting and of short duration. Bolus insulin doses, therefore, are adjusted in real time from meal to meal depending on the blood sugar of the moment. Patients with diabetes needing titration of basal insulin have the option of self-titration [16]. Patients who are not comfortable with self-titration usually return for in-person interactions with their provider. In-person titration requires that frequent appointments be available and that patients are able to attend such appointments.

Safety-net clinics serve a predominantly ethnic minority population who face disproportionate logistical challenges to in-person care (ie, missed-work leading to lost wages, transportation challenges) [12,13,17]. Within this population, remote adjustment of basal insulin for patients with diabetes is an appealing option.

Prior Work

Many cell phone–based programs do exist to help patients track their blood glucose, but we found no programs that only required *basic* cell phone technology to *both* gather blood glucose information and give titration advice [18-25]. To address this, the Mobile Insulin Titration Intervention (MITI) was created [26]. The goal of MITI was to provide a remote basal insulin titration program for patients with type 2 diabetes initiating or titrating basal insulin, utilizing the basic, low-cost, cell phone technology that most patients in safety-net clinics are already using [18-25]: text messages and phone calls.

Patients enrolled in MITI still have to inject insulin and monitor their fasting blood glucose (FBG). Clinicians still need to advise patients on how to adjust their once-daily basal insulin dose. However, with MITI, instead of patients having to travel to the clinic to exchange this information in-person, the clinic comes to the patients through their cell phones.

MITI patients receive a daily weekday text message asking for them to type back their fasting blood sugar and a weekly phone call from a registered nurse advising them on how to adjust their basal insulin dose. Upward titration of basal insulin through the MITI program stops when *one* morning fasting blood sugar level falls into the desired range of 80-130 mg/dL (4.4-7.2 mmol/L), when the patient reaches a basal insulin dose of 50 units, when a patient withdraws from the program, or when 12 weeks elapse. MITI is not designed to be an indefinite program. Rather, it is a (maximum) 12-week opportunity to have one's once-daily basal insulin adjusted in a patient-centered, convenient way. If at the end of 12 weeks, the blood glucose is not controlled, the patient needs to return for in-person evaluation and titration.

MITI was successfully pilot tested in a randomized controlled trial (RCT) with 61 patients and shown to be efficacious at

Bellevue Hospital (the oldest public hospital in the United States) in 2013 and 2014 [26]. In total, 88% percent of pilot study patients (compared with 37% of control patients) were able to find their optimal insulin dose (OID), and they did so in an average of 3 weeks. Additionally, 84% of text message prompts received a response and patients had high satisfaction with the program. On the basis of the success of the pilot study, from April 2016 to April 2017, MITI was implemented into usual care at both Bellevue Hospital and another New York City (NYC) Health and Hospitals site, Gouverneur Health.

A systematic review and network meta-analysis of the comparative effectiveness of different telemedicine strategies was published in October 2017 by Lee et al [27]. The group evaluated 107 telemedicine studies involving over 20,000 patients. Studies focusing on tele-education, telemonitoring, and teleconsultation led to improved HbA_{1c}, with the latter (teleconsultation) being the most effective. Their conclusion was that "assessing the acceptability and implementation challenges of telemedicine in resource-poor areas is an important next step to accelerate translation." This study directly addresses this need for more implementation evaluations in the field of telemedicine.

Goal of This Study

It can take 17 years for clinical interventions to transition from efficacious RCTs to routine clinical practice, and many programs shown to be efficacious during RCTs never make it into usual care [28]. Implementation science studies ways to promote the adoption of programs shown to be effective in RCT into routine care.

The goal of this implementation study was to evaluate MITI as it transitioned from an RCT to real-world settings. Guided by Proctor et al's framework [29] for implementation research, the study assessed key outcomes in 3 domains—implementation (process), service (clinical and program), and client (satisfaction)—to ensure that MITI as usual care continued to be beneficial and acceptable to an expanded underserved population. The study also conducted qualitative interviews with patients and staff during the transition to learn implementation barriers and facilitators, with the goals of making changes in real time to the MITI program based on the feedback that we learned, and of guiding future implementation efforts. This report summarizes the results of the quantitative implementation evaluation. The results of the qualitative interviews learning of barriers to and facilitators of the implementation process are reported separately.

Methods

Study Design

A mixed-methods study was conducted to evaluate the implementation of MITI into routine care at 2 safety-net health care systems in NYC. The study's quantitative evaluation used a single-group, prepost study design to assess the clinical outcomes, core process outcomes, and patient satisfaction with MITI as it became routine care [29]. Our research team conducted in-depth qualitative interviews with patients and staff assessing barriers to and facilitators of the implementation of MITI.

Setting

Bellevue Hospital was the site of the original MITI pilot RCT. At the completion of the pilot program, the RCT data were presented to Bellevue's parent company, NYC Health and Hospitals. NYC Health and Hospitals saw the value in the program and gave approval to transition MITI from a research program to a usual care program available to patients with diabetes needing basal insulin titration in the Bellevue Adult Primary Care Center. For MITI to run effectively, there needed to be a full-time program coordinator, and NYC Health and Hospitals provided this. In 2015, the MITI clinical director (NL) sought to expand MITI to a second site. Gouverneur Health was selected because it is part of NYC Health and Hospitals, an affiliate of NYU, and a clinic that serves the underserved. A common clinical champion made a connection between the MITI program director and the leadership team at Gouverneur Health. Once this connection was made, the Bellevue-based MITI program director met with leadership at Gouverneur Health. The MITI director shared information from the pilot RCT with Gouverneur leadership regarding MITI's efficaciousness, feasibility, adoptability, and patient satisfaction. The Medical Director, Nursing Director, and clinicians at Gouverneur saw the benefit of implementing this efficacious program that had high patient satisfaction and that did not bring a significant additional time burden to staff. Nursing leadership felt that the work of MITI was well within the scope of the general internal medicine clinic nurses at Gouverneur Health.

Thus, between April 2016 and April 2017, MITI was implemented into usual care in the adult medical clinics of Bellevue Hospital and Gouverneur Health in NYC. As stated above, both facilities are members of the NYC Health and Hospital Enterprise (the largest public hospital system in the United States) and affiliates of the New York University School of Medicine. Bellevue Hospital and Gouverneur Health each care for approximately 5000 patients with diabetes. Approximately 30% of patients with diabetes at Bellevue and 10% of patients with diabetes at Gouverneur have an HbA_{1c} of 8% (64 mmol/mol) or above. Both Bellevue and Gouverneur are safety-net centers serving a multiethnic and multiracial patient population. Most patients (65% at Bellevue, 75% at Gouverneur Health) have either Medicaid or are uninsured.

Inclusion and Exclusion Criteria

Patients appropriate for MITI have the following: (1) type 2 diabetes; (2) an HbA_{1c} ≥8% or 64 mmol/mol, (based on the labs obtained at enrollment or the most recent value in the electronic medical record within the 2 months before enrollment); (3) FBG levels in the last 2 weeks ≥130 mg/dL (7.2 mmol/L) but ≤400 mg/dL (22.2 mmol/L); (4) to be starting basal insulin or in need of titration of an existing dose of basal insulin; (5) their basal insulin be either glargine or detemir, dosed once daily; (6) a creatinine ≤1.3 mg/dL (115 μmol/L) for women or ≤1.4 mg/dL (123.8 μmol/L) for men; (7) an age between 18 and 70 years (during this study's evaluation period, an exception was made for 1 patient above the age of 70 years. The patient was inadvertently enrolled by one of the team nurses and it was not

brought to the MITI team's attention until mid-way through the program.); (8) a cell phone; and (9) the ability to receive text messages, to understand the text messages in either English or Spanish, to check their home FBG, to text back the results, and to accept and make phone calls.

Patients were excluded if they did not meet the above inclusion criteria, if they were taking oral steroids at the time of enrollment, if they had taken oral/injectable steroids within the past 2 weeks, or if they were currently taking rapid-acting insulin.

Any patients referred to MITI who had not injected insulin previously were sent for education on how to do so before enrollment.

Mobile Insulin Titration Intervention Program Referral, Enrollment, and Consent

Providers and nurses at the 2 clinics were educated about the MITI program before implementation. When a patient needed once-daily basal insulin titration, and preferred to receive titration instructions remotely versus coming back to the clinic, providers referred the patient to MITI. Patients were almost always enrolled immediately following an in-person visit. However, there were a few times when the MITI program was notified of a referral for a patient that was not currently in clinic, and program staff arranged for the patient to return for enrollment.

At enrollment, a MITI team member met with the patient to confirm eligibility, administer a survey capturing demographics, explain the program in more detail, and enroll the patient into a secure Web-based messaging platform developed and maintained by Wellpass (Wellpass, Inc, New York, NY) [30]. Messages sent as short message service (SMS) texts from the Wellpass platform are not encrypted. Therefore, patients at enrollment also signed a consent form that gave their written authorization to exchange protected health information (eg, FBG levels) via text.

Intervention: Text Messages, Monitoring, Titration, and Goals

Every weekday at a patient-specified time, MITI participants received a text message asking, "What was your fasting blood sugar this morning?" Responses were monitored daily for any "alarm" values, defined as FBG <80 mg/dL [, or 4.4 mmol/L, or FBG >400 mg/dL, or 22.2 mmol/L, which were addressed by the monitoring nurses in real time. Registered nurses called all patients once weekly to advise on dose titration using the structured algorithm that the MITI team developed before and tested during the 2013 pilot study [26].

The MITI coordinator also monitored for the presence of daily responses. If patients did not respond with their FBG value 3 days in a row, the MITI program coordinator reached out to the patient via phone calls to understand the reason for the lack of the text message response and problem solve, if needed.

The goal of the program is to find the patient's OID. OID is defined as the once-daily basal insulin dose that leads to one FBG of 80-130 mg/dL (4.4-7.2 mmol/L) inclusive or that reaches the 50 unit maximum dose. The program ends the week that the patient reaches their OID, when 12 weeks elapse, or when the program terminates early (ie, when a patient actively withdraws from the program, becomes lost to follow-up, or when patient eligibility changes, as occurred with one patient whose doctor decided soon after referral to MITI that rapid-acting insulin should be started). Patients were considered lost to follow-up when the MITI nurse was not able to reach them despite efforts over 3 consecutive weeks.

Two Different Staffing Models

The 2 sites used different staffing models to serve the following 2 functions: (1) enrolling, and (2) monitoring and titrating. At Bellevue Hospital, the enrollment onto the secure Web platform and into the MITI program was carried out by the on-site MITI program coordinator. The daily monitoring for alarm values and weekly phone calls at Bellevue Hospital were primarily carried out by 1 of the clinic's 2 diabetes nurses (both are certified diabetes educators and registered nurses). The diabetic nurses had structured time set aside each week to carry out the work of the MITI program.

At Gouverneur Health, there is no full-time onsite MITI coordinator or diabetes nurse. The medicine clinic nurses enrolled patients into MITI as part of the routine clinic discharge process. At Gouverneur Health, these same medicine nurses were also the team members responsible for checking alarm values daily and making titration calls weekly. This monitoring and titration work was carried out as part of the daily workflow.

Mobile Insulin Titration Intervention Implementation Process

Evidence-based strategies to implement MITI were used at each site [31]. For example, before the rollout of MITI as usual care, a multidisciplinary advisory board was formed at each site to guide implementation. Potential referring providers and MITI nurses were trained about eligibility criteria, enrollment procedures, daily monitoring, and weekly titration calls. MITI nurses attended multiple in-person trainings on how to use the intervention software (eg, logging on to the system, enrolling patients, monitoring daily text messages). The MITI team attended routine staff meetings after the rollout to give updates and get feedback. As new providers and nurses joined the clinics, they were educated individually. The MITI coordinator was available on an ongoing basis to troubleshoot any issues that arose, review standard procedures, answer questions, and provide information on any updates and/or changes to the program.

Measures and Outcomes

The study measured multiple program outcomes, clinical outcomes, process outcomes, and satisfaction at both sites informed by the Proctor et al model of outcomes in implementation research [29] to study implementation processes and study outcomes in 2 different clinical settings.

Mobile Insulin Titration Intervention Program Outcomes

We measured the percentage of patients who achieved their OID and the number of days required to reach OID. We evaluated the frequency of each component of OID (reaching a FBG of 80-130 mg/dL [4.4-7.2 mmol/L] inclusive or reaching

the maximum dose of 50 units). We evaluated the percentage of patients that did not reach OID by 12 weeks or for whom the program was terminated early (ie, no longer met eligibility criteria, withdrew from the program, or were lost to follow-up).

Mobile Insulin Titration Intervention Cost Analysis

An analysis of potential cost savings was conducted. The cost of the time of nursing, administration, and the MITI program director, calculated from national wage data obtained from the Bureau of Labor Statistics [32], in addition to technology costs (both technology setup fees and individual staff licenses to access the Web platform [30]), were calculated. Savings associated with patient time (also calculated from national wage tables from the Bureau of Labor Statistics [32]) and with the averted in-person medical clinic visit (based on Healthcare Bluebook fee scales [33]) were evaluated.

Mobile Insulin Titration Intervention Clinical Outcomes

Additional clinical measures included FBG values on the first day of MITI as well as the day that qualified the patient for completion, rates of hypoglycemia, and, when available, follow-up HbA_{1c} values were abstracted from the patient's medical record.

Pre-MITI HbA_{1c} blood test values were accepted when obtained within 2 months of enrollment, and post-MITI HbA_{1c} values were accepted between 2 and 6 months after completion of MITI. Ideally, the MITI HbA_{1c} values would be obtained on the day of enrollment and then once again 2-3 months after program termination. However, in real-world settings (and not RCTs), it is the referring clinician who decides if new labs are needed on the day of enrollment or if the HbA_{1c} from the recent past still represents the glycemic control range for the patient in question. Likewise, for the post-MITI HbA_{1c} s, it was not within the scope of the program to bring patients back in for a special blood test that best fit our time frame. Referring providers were in charge of ordering the post-MITI HbA_{1c} blood tests and, although ideally we would have liked to have seen them between 2-3 months after program completion, we accepted values up to 6 months if this was the time frame within which patients were able to return.

Mobile Insulin Titration Intervention Process Outcomes

MITI process outcomes were assessed using data collected by the MITI program as part of routine operations. The Wellpass system captures all texts sent and received, which we used to calculate the percentage of texts prompts that received a patient response. We used clinic administrative data to calculate MITI uptake by providers, defined as the percent of providers who referred at least one patient to MITI, as well as to calculate the percentage of weeks that the nurses were able to connect with the patients to provide titration instructions. MITI nurses documented the time it took per patient per week to carry out the titration intervention.

Patient Satisfaction and Patient Time Saved

Patients completed a short survey at enrollment, which asked them how long it took them to travel to the clinic and how long they waited for their appointment, which we used to calculate patient time saved when an in-person visit was averted. The MITI coordinator at Bellevue called all MITI patients following completion of the program to assess patient satisfaction with the program. Patient satisfaction surveys evaluated 6 factors using a Likert-type scale: comfort level sharing information through text, preference for clinic, thoughts on the number of text messages (eg, too few, too many, just right), text helpfulness as a reminder to check sugar levels, text helpfulness as a reminder to take insulin, and text helpfulness as a reminder to make healthy food choices. Patients were not compensated for their participation.

Statistical Analysis

Descriptive statistics were used to summarize sample demographics and quantitative outcomes, stratified by site. Paired *t* tests were used to examine whether there were significant within-group changes in HbA_{1c} and FBG for the MITI group between baseline and at the end of MITI program. A two-sided $P<.05$ was considered to be statistically significant, and all analyses were performed in SAS 9.4 (SAS Institute Inc, Cary, NC). For the cost analysis, the total costs for each patient were subtracted from the total savings for each patient. These net savings are presented in terms of "per-patient per-week." To understand net savings on a larger scale, we took these "per-patient per-week" savings and calculated how they would change based on various projections of how many patients would utilize the program in 1 year.

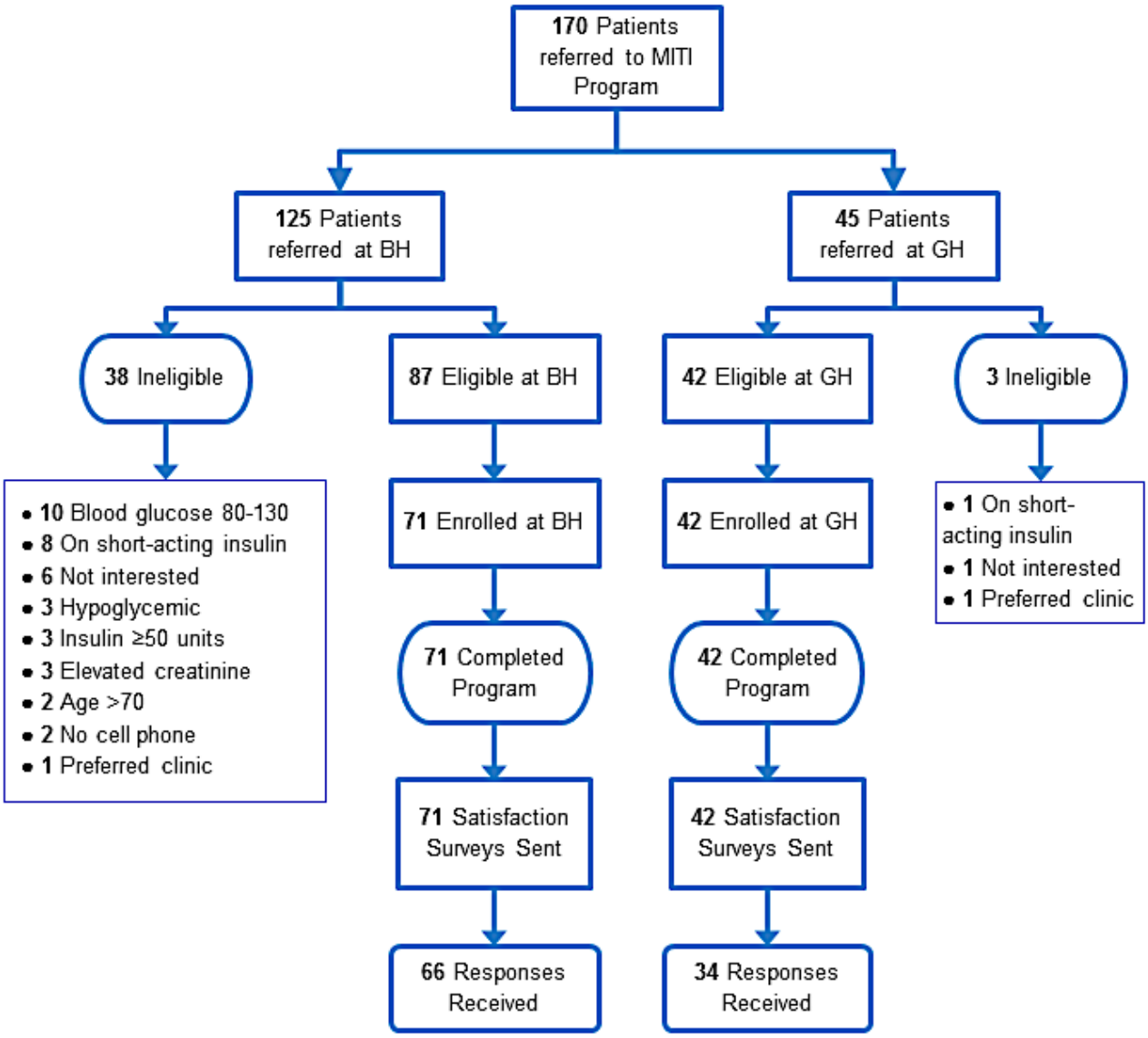
Results

Patient Characteristics

Figure 1 displays patient flow through MITI. Across the 2 sites, providers referred 170 patients to the MITI program—129 of whom (75.9%) were eligible. Of these, 113 (87.6%) enrolled. A total of 41 patients were excluded for the following reasons: 10 had a FBG between 80-130 mg/dL (4.4-7.2 mmol/L) in the past 2 weeks, 9 were on short-acting insulin, 3 had any blood glucose (fasting or non-fasting) <80 mg/dL (4.4 mmol/L) in the past 2 weeks, 3 were already taking ≥50 units of insulin, 3 had elevated creatinine, 2 preferred clinics, 2 were older than 70 years, 2 did not have a cell phone, and 7 declined to participate for other reasons. Of the remaining 16 patient referrals, 15 were lost to follow-up and 1 was enrolled in MITI's year 2 (Figure 1).

Table 1 shows MITI patient demographics. MITI patients had an average age of 50 years (standard deviation, SD=10), 45.1% were female, 78.8% Hispanic, 42.5% unemployed, and 46.0% uninsured. Moreover, 59% chose Spanish as their text language and 40.7% chose English. The initial FBG value was 209 mg/dL (11.6 mmol/L, SD 71) and the mean initial HbA_{1c} level was 11.4% (101 mmol/mol, SD 1.9%). Patients spent on average 142 min (SD 69) traveling to and from the clinic and waiting in the waiting room (52 min travel time each way, SD 27, 43 min waiting room time, SD 39).

Figure 1. Mobile insulin titration intervention (MITI) program referrals and enrollment diagram. BH: Bellevue Hospital; GH: Gouverneur Health.



Mobile Insulin Titration Intervention Cost Analysis

A cost analysis indicated that the value of patients' time and the cost of the averted clinic visits outweighed the cost of the MITI program. The per-patient per-week savings at Bellevue Hospital ranged from US $169.65, if 71 patients per year (the actual number of patients that participated in MITI during this study) participated in the program, to US $176.91 if 100 patients per year participated in the program, to US $185.80 if 200 patients per year participated in the program. At Gouverneur Health, the per-patient per-week savings ranged from US $0.94, if 42 patients participated in the program per year (the actual number of patients that participated in MITI during this study), to US $107.55 if 100 patients participated in the program per year, to US $146.15 if 200 patients participated in the program per year.

Mobile Insulin Titration Intervention Program Outcomes

Table 2 shows MITI program outcomes related to the frequency of each program discharge status (eg, optimal insulin dose, 12 weeks without optimal insulin dose, program terminated early). In total, 84.1% of patients reached their optimal insulin dose, and they did so in an average of 24 days. This included 74.3% of patients who reached a once-daily basal insulin dose to bring one blood glucose down to between 80 and 130 mg/dL (4.4-7.2 mmol/L), inclusive, as well as an additional 9.7% of patients who reached the maximum dose of 50 units. Moreover, 8.0% of patients did not reach the goal by 12 weeks and 8.0% of patients terminated the program early (1 patient had short-acting insulin added to their regimen and thus became ineligible, 1 decided she did not want to take insulin, 1 patient actively withdrew, and 6 were lost to follow-up and could not be reached despite multiple attempts over 3 weeks).

Table 1. Enrolled Mobile Insulin Titration Intervention (MITI) patient demographics.

Demographics	Bellevue (N=71)	Gouverneur (N=42)	Total (N=113)
Age			
Mean age (SD; min-max)	50 (10; 26-69)	50 (11; 24-73)	50 (10; 24-73)
Female gender, n (%)	34 (48)	17 (41)	51 (45.1)
Race, n (%)			
Native Hawaiian or Pacific Islander	0 (0.0)	0 (0.0)	0 (0.0)
Native American or Alaskan Native	1 (1)	1 (2)	2 (1.8)
Asian	5 (7.0)	3 (7)	8 (7.1)
White	7 (10)	3 (7)	10 (8.8)
Black or African American	8 (11.3)	3 (7)	11 (9.7)
Other[a]	50 (70)	32 (76)	82 (72.6)
Hispanic ethnicity[a], n (%)	54 (76)	35 (83)	89 (78.8)
No employment, n (%)	36 (51)	12 (29)	48 (42.5)
No health insurance, n (%)	39 (55)	13 (31)	52 (46.0)
Had a visit co-payment? Yes, n (%)	34 (489)	14 (33)	48 (42.5)
Language used for text messages, n (%)			
Spanish	41 (58)	26 (62)	67 (59.3)
Pre-MITI patient travel and wait time in min, mean (SD)	157 (66)	117 (66)	142 (69)

[a]All patients who checked the "Other" race option said they were Hispanic.

Table 2. Mobile Insulin Titration Intervention (MITI) program outcomes.

MITI discharge status	Bellevue (N=71)			Gouverneur (N=42)			Total (N=113)		
	n (%)	Mean days in MITI (min-max)	SD	n (%)	Mean days in MITI (min-max)	SD	n (%)	Mean days in MITI (min-max)	SD
Achieved optimal insulin dose	57 (80)	26 (3-56)	24	38 (91)	21 (3-77)	22	95 (84.1)	24 (2-84)	23
Reached 80-130	50 (70)	21 (3-84)	20	34 (81)	17 (2-42)	17	84 (74.3)	20 (2-84)	19
Reached max insulin dose (50 units)	7 (10)	62 (27-84)	22	4 (10)	57 (10-77)	32	11 (9.7)	60 (10-84)	25
Not at goal by 12 weeks	7 (10)	84	N/A[a]	2 (5)	84	N/A	9 (8.0)	84	N/A
Program terminated early	7 (10)	35 (16-63)	18	2 (5)	37 (16-58)	30	9 (8.0)	36 (16-63)	19

[a]N/A: not applicable.

Table 3. Mobile insulin titration intervention (MITI) clinical outcomes for fasting blood glucose values.

MITI clinical outcomes	Bellevue (N=70[a])	Gouverneur (N=42)	Total (N=112[a])
Mean fasting blood glucose value			
First day of MITI in mg/dL (SD), mmol/L (SD)	214 (73), 11.9 (4.2)	201 (68), 11.2 (4.1)	209 (71), 11.6 (4.2)
Last day of MITI in mg/dL (SD), mmol/L (SD)	144 (48), 8.0 (2.9)	136 (41), 7.6 (2.8)	141 (45), 7.8 (3.0)
P value	<.001	<.001	<.001

[a]There are missing data for one Bellevue patient who never texted.

Table 4. Mobile insulin titration intervention (MITI) clinical outcomes for HbA_{1c} values.

MITI clinical outcomes	Bellevue (N=51[a])	Gouverneur (N=29[a])	Total (N=80[a])
Mean pre-MITI A_{1c} in % (SD), mmol/mol (SD)	11.6 (1.8), 103 (20)	11.1 (2.0), 98 (22)	11.4 (1.9), 101 (21)
Mean post-MITI A_{1c} in % (SD), mmol/mol (SD)	10.3 (2.1), 89 (23)	9.4 (2.2), 79 (24)	10.0 (2.2), 86 (24)
P value	<.001	<.003	<.001

[a]Data are only included for those patients who have both a pre- and post-HbA_{1c}.

Mobile Insulin Titration Intervention Clinical Outcomes

Table 3 shows MITI clinical outcomes related to changes in fasting blood glucose values. The mean FBG on the first day of the program was 209 mg/dL (11.6 mmol/L, SD 71) and it fell to 141 mg/dL (7.8 mmol/L, SD 45) on the day that qualified the patient for completion of the program, *P*<.001. Of note, out of 2049 texted blood glucose values (and any additional information shared during the weekly phone calls with the MITI nurses), there were only 2 reports of hypoglycemia; neither were severe.

Moreover, 80 of the 113 MITI patients have had HbA_{1c} values within both windows for pre and post lab collection and were included in our main HbA_{1c} analysis (Table 4).

The breakdown of when the 80 pre HbA_{1c} lab values were collected is as follows: 20 on the day of enrollment, an additional 22 were collected within 1 week before enrollment, another 11 were collected within 2 weeks of enrollment, and the remaining 27 within 2 months of enrollment. The average time of pre HbA_{1c} lab draw was within 17.2 days of enrollment. Post HbA_{1c} lab collection had to be within 6 months of program completion date. The review found that 34 of the 80 patients had their post HbA_{1c} labs collected within 3 months of program completion. The remaining 46 had labs collected by the end of the 6-month window. The average time of post HbA_{1c} lab collection was at 104.1 days after program completion.

HbA_{1c} levels fell from a mean of 11.4% (101 mmol/mol, SD 1.9%) at enrollment (SD 21) to 10.0% (86 mmol/mol, SD 2.2%) at follow-up (SD 24), *P*=.003. Moreover, 43 of the 80 patients had an HbA_{1c} that fell by at least 1% point. In addition, 29 of the 80 patients had an HbA_{1c} that fell by ≥2% points. Furthermore, 17 the 80 patients with HbA_{1c} results have an HbA_{1c} that fell to ≤8% (64 mmol/mol).

Of note, the pre HbA_{1c} mean for the 96 of 113 enrollees who had a baseline HbA_{1c} within 2 months of enrollment (not just of the 80 that had *both* pre and post values) had essentially the same mean of 11.6% (103 mmol/mol), supporting that those that have had follow-up data available in the electronic medical record were representative of the entire population in terms of disease severity at the start of MITI.

Mobile Insulin Titration Intervention Process Outcomes

Table 5 shows MITI process outcomes. We found that 90.1% of MITI's text message prompts received a response, demonstrating a very high acceptability and participation rate for those enrolled in the MITI program. Review of the data showed that nurses were able to connect with patients to provide titration instructions 81.9% of the time. Median nurse time was 15 min per patient per week to carry out the titration intervention, both at Bellevue Hospital and Gouverneur Health. This includes the time required to prepare for the weekly titration phone call, make the phone call (which often included an interpreter being on the line), and to document the phone call.

When examining the proportion of providers who made at least one referral to MITI, we found that 83% of attending physicians, 42% of resident physicians, 62% of physician assistants, and 100% of both diabetes educators at Bellevue made at least one referral (Table 6)

Patient Satisfaction Outcomes

Table 7 shows patient satisfaction data. Patient satisfaction surveys revealed that 97.0% of patients were comfortable sharing their information through text, 97.0% preferred not having to come to clinic, and 94.0% thought that the number of texts received were just right (of note, the other 6% thought they were too few). Surveys also showed that 96.0% felt that the texts were somewhat or very helpful as reminders to check their blood glucose, 84.0% felt the texts were somewhat or very helpful as reminders to take their insulin, and 87.0% felt the texts were somewhat or very helpful as reminders to make healthy food choices.

Table 5. Mobile insulin titration intervention (MITI) process outcomes.

MITI process outcomes	Bellevue	Gouverneur	Total
Text response rate (responses ÷ prompts), n (%)	1387/1530 (90.65)	662/744 (89.0)	2049/2274 (90.11)
Call connection rate (connections ÷ weeks), n (%)	264/340 (77.6)	152/168 (90.5)	416/508 (81.9)
Median nurse time (min) for weekly titration interaction[a] (per-patient per-week), mean (SD)	15 (7)	15 (6)	15 (7)

[a]Titration Interaction time includes the time to prepare for the call, have the call (often with a translator), and document the call.

Table 6. Percentage of providers making at least one referral to the Mobile Insulin Titration Intervention (MITI) program.

Provider type	Bellevue, n (%)[a]	Gouverneur, n (%)[a]	Total, n (%)[a]
Attending physician	28 (97)	13 (68)	41 (85)
Resident	37 (40)	4 (15)	39[b] (42)
Physician assistant	3 (50)	5 (71)	8 (62)
Diabetes nurses	2 (100)	N/A[c]	2 (100)

[a]Referral rates calculated as the number of unique referring clinicians divided by the total number of possible referring clinicians.

[b]A small number of residents rotated through and made referrals at both sites during the study period.

[c]N/A: not applicable.

Table 7. Mobile insulin titration intervention (MITI) patient satisfaction and program feedback among patients who have completed the program.

Patient experience questions	Bellevue (N=66 survey respondents)	Gouverneur (N=34 survey respondents)	Total (N=100 survey respondents)
Comfort level sharing info via texts, n (%)			
Very	65 (99)	32 (94)	97 (97.0)
Somewhat	0 (0)	2 (6)	2 (2.0)
Not at all	1 (2)	0 (0)	1 (1.0)
Preference for clinic, n (%)			
No	65 (99)	32 (94.)	97 (97.0)
Yes	1 (2)	2 (6)	3 (3.0)
Number of texts were, n (%)			
Too many	0 (0)	0 (0)	0 (0.0)
Just right	64 (97)	30 (88)	94 (94.0)
Too few	2 (3)	4 (12)	6 (6.0)
Texts helpfulness as reminder to check sugar levels, n (%)			
Very	48 (72)	18 (53)	66 (66.0)
Somewhat	16 (24)	14 (41)	30 (30.0)
Not at all	2 (3)	2 (6)	4 (4.0)
Texts helpfulness as reminder to take insulin[a], n (%)			
Very	24 (36)	10 (29)	34 (34.0)
Somewhat	33 (50)	17 (50)	50 (50.0)
Not at all	9 (14)	7 (21)	16 (16.0)
Texts helpfulness as reminder to make healthy food choices, n (%)			
Yes, often	39 (59)	16 (47)	55 (55.0)
Sometimes	20 (30)	12 (35)	32 (32.0)
Never	7 (11)	6 (18)	13 (13.0)

[a]Patients who said that the texts were not helpful as a reminder to take insulin said that they were already accustomed to taking their insulin each day before the program.

Technology

Occasionally during the course of the program, there was a random day where the text messages were not delivered. The MITI program coordinator (who checks daily to make sure the program is running) noted this when it occurred and reached out to Wellpass to troubleshoot the issues. It was always remedied quickly.

Discussion

Principal Findings

This study evaluated the implementation of a mobile intervention to titrate basal insulin for uncontrolled type 2 diabetes patients (“MITI”), as it became usual care at 2 ambulatory clinics in NYC. The study produced several important findings that help address existing literature gaps

regarding the real-world implementation of telemedicine interventions, particularly in resource-poor settings such as US safety-net clinics [27]. A strength of the study design was its evaluation of MITI as it transitioned into usual care at 2 sites that had different patient populations and very different staffing models. Bellevue Hospital used a designated program coordinator to enroll patients into the MITI program and then specialty trained diabetes nurses to carry out MITI's clinical work (daily monitoring and weekly titration phone calls). Gouverneur Health used the general medical clinic registered nurses to both enroll patients (as part of the routine clinic outtake process) and to carry out MITI's clinical work. Both sites found success, showing MITI to be adaptable and flexible in different settings. The latter setting, in particular, highlighted MITI as generalizable to general medical clinics where registered nurses (but perhaps not program coordinators nor specialty trained diabetes nurses) will be present.

Program and clinical data (ie, percent of patients reaching OID, time to OID, changes in FBG and HbA_{1c} levels) showed that patients enrolled in the MITI program achieved excellent clinical outcomes that are very similar to those found in MITI's prior pilot RCT [26]. This demonstrates that MITI was able to achieve continued positive outcomes in real-world settings using regular clinical staff serving expanded safety net populations. Moreover, the study of a patient population receiving MITI as usual care shows that MITI is able to produce positive outcomes with patients who did not self-select to participate in a controlled research study.

Process outcomes (ie, text and titration call response rates) further demonstrated high patient engagement, acceptability, and usability of the program among patients. Previous studies have reported that diabetes patients find telehealth interventions easy to use and that they engage with text-messaging interventions, and prior implementation research has shown that employing user-friendly technology that takes little time to learn and use in telemedicine interventions for diabetes patients is a key component of implementation success [34]. In this study, patients were easily able to use MITI's once-daily basic SMS messaging and brief weekly phone calls. Additionally, satisfaction survey responses found that most patients were satisfied with the number of texts received, were comfortable sharing their information through text, and preferred not having to come to the clinic. Although the MITI daily text only included the question, "What was your fasting blood sugar this morning?", without additional motivational or educational content, patients repeatedly shared on surveys that just knowing that the text message was coming in the morning provided the motivation to make healthier food choices, take their insulin, and check their blood glucose in the morning. Patient satisfaction ratings with MITI were similar to—or greater than—prior research evaluating telemedicine interventions for diabetes patients [35-37]. For example, Odnoletkova et al found that 98% of patients enrolled in a telecoaching intervention for diabetes patients reported overall satisfaction with the program and 92% felt that phone was an acceptable mode of communication [35]. Welch and Balder found high acceptability, convenience, and ease of use ratings (>80%) from 30 diabetes patients enrolled in a multi-component telehealth program being studied at an urban community center. Of note, prior reported data on the acceptability of telemedicine interventions for diabetes patients have largely been collected during feasibility or controlled studies [37]. This implementation study makes a significant contribution to the literature on telemedicine interventions for diabetes patients by demonstrating high patient acceptability and usability of MITI in a real-world, safety-net setting.

Low adoption by providers can be one of the greatest barriers to implementation of telehealth interventions, yet our referral data showed high adoption of MITI by staff, with most providers making at least one MITI referral. This report did not assess acceptability among providers. Detailed information about acceptability will be reported in a separate manuscript. Briefly, interviewed providers who made at least one referral reported that their perception of MITI as an effective, convenient, usable, and acceptable option for their patients drove their adoption of the intervention (consistent with other implementation studies evaluating factors that influence provider use of new telehealth programs) [35,38,39]. Interviewed providers who did not refer a single patient provided reasons such as patients neither English- nor Spanish-speaking (the only two languages our program was able to offer) or older and less technologically "savvy." Prior implementation research has also suggested that providers may not refer patients to mHealth interventions because of their concerns about the effectiveness of mobile health care as compared with direct patient contact [34].

Finally, process outcomes showed the program to be feasible as observed nurse titration time of 15 min per patient per week was reasonable. The study's cost-savings analysis further showed that the MITI program's savings in terms of visits averted and patient time saved far outweighed the costs of technology and staff time at both sites. Of note, the per-patient per-week savings are less at Gouverneur Health because the technology company that we worked with charges a license for each clinician that accesses the platform. At Bellevue Hospital, there were only 5 team members that needed to access the platform. These 5 team members included the 2 diabetes nurses who ran the program for all MITI patients, 1 general medical clinic nurse (who was a backup in case both diabetes nurses were away), and the MITI clinical coordinator and program director (both of whom oversaw the program at both locations, but spent the majority of their combined time at Bellevue Hospital). In contrast, at Gouverneur Health, there were 18 nurses that participated in the MITI program, and each of them needed their own license. Of note, this cost sensitivity analysis is based on the pricing of one Web platform provider. In addition, this cost analysis does not take into account potential downstream health savings from improved glycemic control.

Limitations

Similar to many implementation research studies, our study had no control group for the clinical outcomes. However, the original MITI pilot RCT showed that MITI was efficacious compared with usual care [26]. Nurse time spent on the project was based on self-report, which may be subject to recall bias. Moreover, we only had pre and post HbA_{1c} data on 71% of MITI patients. It is possible that there was a clinical difference between these

patients, and the 29% of patients who did not have pre and post labs within the lab window. Additionally, a further limitation of this study might be our gap of knowledge in the potential reach of MITI. It is hard for us to know how many patients would have been eligible for the program, as there is no registry of such patients.

Conclusions

MITI is a patient-centered, text message-based program that allows the remote titration of once-daily basal insulin solely through the use of basic cell phone technology, which has great potential to improve access to care and reduce disparities in diabetes care for a multiracial, multiethnic, low-income population. This implementation study showed MITI to have continued success after transitioning from an RCT pilot program into real-world settings. MITI was found to flexible in different settings, generalizable to a general medical clinic setting, highly acceptable to patients and providers, and feasible for nurses to deliver as part of their routine workflow.

Acknowledgments

This study was supported by the Aetna Foundation through a Healthier World Challenge Innovation Grant (Grant Number 15-1470, PI: NKL). The sponsor did not participate in any aspect of the program design, data collection, data analysis, or manuscript development. The authors wish to thank Mr Yiding Jiang for his early contribution to data analysis. The authors wish to thank the MITI patients and MITI nurses, who participated in our program and without whom this work would not have been possible. The authors would also like to acknowledge the staff of the Adult Primary Care Centers at Bellevue Hospital and Gouverneur Health, as well as New York City Health and Hospitals, who gave tremendous support to the program

Authors' Contributions

NL conceptualized and designed the study, interpreted the results, wrote the manuscript, and reviewed/edited the manuscript. NOB helped with data collection and reviewed/edited the manuscript; SRA and DNM conducted qualitative interviews and reviewed/edited the manuscript. LH, ATL, and MAS conceptualized and designed the study, interpreted the results, and reviewed/edited the manuscript. LN interpreted results and reviewed/edited the manuscript. AP analyzed the cost sensitivity of the program and reviewed/edited the manuscript. BW analyzed the data, interpreted the results, and reviewed/edited manuscript. ESR conceptualized and designed the study, conducted qualitative interviews, interpreted the results, and reviewed/edited the manuscript. NL is the guarantor and takes responsibility for the contents of the manuscript.

Conflicts of Interest

None declared.

References

1. Centers for Disease Control and Prevention. 2017. National diabetes statistics report: estimates of diabetes and its burden in the United States URL: https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf [WebCite Cache ID 6v1ZDSFYb]
2. Boyle JP, Thompson TJ, Gregg EW, Barker LE, Williamson DF. Projection of the year 2050 burden of diabetes in the US adult population: dynamic modeling of incidence, mortality, and prediabetes prevalence. Popul Health Metr 2010 Oct 22;8:29 [FREE Full text] [doi: 10.1186/1478-7954-8-29] [Medline: 20969750]
3. Dieleman JL, Baral R, Birger M, Bui AL, Bulchis A, Chapin A, et al. US spending on personal health care and public health, 1996-2013. J Am Med Assoc 2016 Dec 27;316(24):2627-2646 [FREE Full text] [doi: 10.1001/jama.2016.16885] [Medline: 28027366]
4. American Diabetes Association. Economic costs of diabetes in the U.S. in 2012. Diabetes Care 2013 Apr;36(4):1033-1046 [FREE Full text] [doi: 10.2337/dc12-2625] [Medline: 23468086]
5. Emerging Risk Factors Collaboration, Sarwar N, Gao P, Seshasai SR, Gobin R, Kaptoge S, et al. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. Lancet 2010 Jun 26;375(9733):2215-2222 [FREE Full text] [doi: 10.1016/S0140-6736(10)60484-9] [Medline: 20609967]
6. Ismail-Beigi F. Glycemic management of type 2 diabetes mellitus. N Engl J Med 2012 Jul 12;367(2):181-183. [doi: 10.1056/nejmc1205238]
7. King P, Peacock I, Donnelly R. The UK prospective diabetes study (UKPDS): clinical and therapeutic implications for type 2 diabetes. Br J Clin Pharmacol 1999 Nov;48(5):643-648 [FREE Full text] [Medline: 10594464]
8. Adler AI, Boyko EJ, Ahroni JH, Stensel V, Forsberg RC, Smith DG. Risk factors for diabetic peripheral sensory neuropathy. Results of the Seattle Prospective Diabetic Foot Study. Diabetes Care 1997 Jul;20(7):1162-1167. [Medline: 9203456]
9. Klein R. Hyperglycemia and microvascular and macrovascular disease in diabetes. Diabetes Care 1995 Feb;18(2):258-268. [Medline: 7729308]
10. Stratton IM, Adler AI, Neil HA, Matthews DR, Manley SE, Cull CA, et al. Association of glycaemia with macrovascular and microvascular complications of type 2 diabetes (UKPDS 35): prospective observational study. Br Med J 2000 Aug 12;321(7258):405-412 [FREE Full text] [Medline: 10938048]

11. Klonoff DC. United Kingdom prospective diabetes study follow-up studies establish a legacy effect of therapy for hyperglycemia but not hypertension. J Diabetes Sci Technol 2008 Nov;2(6):922-924 [FREE Full text] [doi: 10.1177/193229680800200601] [Medline: 19885280]
12. AHRQ. Disparities in health care quality among racial and ethnic minority groups: selected findings from the AHRQ 2010 NHQR and NHDR URL: https://www.ahrq.gov/professionals/quality-patient-safety/index.html [accessed 2017-11-16] [WebCite Cache ID 6v1ZG24H3]
13. Chow EA, Foster H, Gonzalez V, McIver L. The disparate impact of diabetes on racial/ethnic minority populations. Clin Diabetes 2012 Jul 16;30(3):130-133. [doi: 10.2337/diaclin.30.3.130]
14. Centers for Disease Control and Prevention. 2011. National diabetes fact sheet: national estimates and general information on diabetes and prediabetes in the United States URL: https://www.cdc.gov/diabetes/pubs/pdf/ndfs_2011.pdf [WebCite Cache ID 6v1aid4q6]
15. American Diabetes Association. Pharmacologic approaches to glycemic treatment. Diabetes Care 2017 Jan;40(Suppl 1):S64-S74. [doi: 10.2337/dc17-S011] [Medline: 27979895]
16. Blonde L, Merilainen M, Karwe V, Raskin P, TITRATE Study Group. Patient-directed titration for achieving glycaemic goals using a once-daily basal insulin analogue: an assessment of two different fasting plasma glucose targets - the TITRATE study. Diabetes Obes Metab 2009 Jun;11(6):623-631. [doi: 10.1111/j.1463-1326.2009.01060.x] [Medline: 19515182]
17. Eisert SL, Mehler PS, Gabow PA. Can America's urban safety net systems be a solution to unequal treatment? J Urban Health 2008 Sep;85(5):766-778 [FREE Full text] [doi: 10.1007/s11524-008-9296-5] [Medline: 18553134]
18. Hsu WC, Lau KH, Huang R, Ghiloni S, Le H, Gilroy S, et al. Utilization of a cloud-based diabetes management program for insulin initiation and titration enables collaborative decision making between healthcare providers and patients. Diabetes Technol Ther 2016 Feb;18(2):59-67 [FREE Full text] [doi: 10.1089/dia.2015.0160] [Medline: 26645932]
19. Pulizzi J, Bollyky J, Toles A, Boulos M, Schneider J. Targeted diabetes education text messaging program increases requests for certified diabetes educator coaching and improves blood glucose trends. J Med Internet Res 2016 Dec 30;2(1):e32. [doi: 10.2196/iproc.6163]
20. Nelson LA, Mulvaney SA, Gebretsadik T, Johnson KB, Osborn CY. The MEssaging for Diabetes (MED) intervention improves short-term medication adherence among low-income adults with type 2 diabetes. J Behav Med 2016 Dec;39(6):995-1000. [doi: 10.1007/s10865-016-9774-2] [Medline: 27488604]
21. Nelson LA, Mulvaney SA, Johnson KB, Osborn CY. mHealth intervention elements and user characteristics determine utility: a mixed-methods analysis. Diabetes Technol Ther 2017 Jan;19(1):9-17. [doi: 10.1089/dia.2016.0294] [Medline: 28099052]
22. Fatehi F, Gray LC, Russell AW. Mobile health (mHealth) for diabetes care: opportunities and challenges. Diabetes Technol Ther 2017 Jan;19(1):1-3. [doi: 10.1089/dia.2016.0430] [Medline: 28099051]
23. Wei NJ, Nathan DM, Wexler DJ. Glycemic control after hospital discharge in insulin-treated type 2 diabetes: a randomized pilot study of daily remote glucose monitoring. Endocr Pract 2015 Feb;21(2):115-121 [FREE Full text] [doi: 10.4158/EP14134.OR] [Medline: 25148814]
24. Nundy S, Dick JJ, Chou C, Nocon RS, Chin MH, Peek ME. Mobile phone diabetes project led to improved glycemic control and net savings for Chicago plan participants. Health Aff (Millwood) 2014 Feb;33(2):265-272 [FREE Full text] [doi: 10.1377/hlthaff.2013.0589] [Medline: 24493770]
25. Arora S, Peters AL, Burner E, Lam CN, Menchine M. Trial to examine text message-based mHealth in emergency department patients with diabetes (TExT-MED): a randomized controlled trial. Ann Emerg Med 2014 Jun;63(6):745-54.e6. [doi: 10.1016/j.annemergmed.2013.10.012] [Medline: 24225332]
26. Levy N, Moynihan V, Nilo A, Singer K, Bernik LS, Etiebet M, et al. The mobile insulin titration intervention (MITI) for insulin glargine titration in an urban, low-income population: randomized controlled trial protocol. JMIR Res Protoc 2015;4(1):e31 [FREE Full text] [doi: 10.2196/resprot.4206] [Medline: 25794243]
27. Lee SW, Chan CK, Chua SS, Chaiyakunapruk N. Comparative effectiveness of telemedicine strategies on type 2 diabetes management: a systematic review and network meta-analysis. Sci Rep 2017 Oct 04;7(1):12680. [doi: 10.1038/s41598-017-12987-z] [Medline: 28978949]
28. Westfall JM, Mold J, Fagnan L. Practice-based research--"Blue Highways" on the NIH roadmap. J Am Med Assoc 2007 Jan 24;297(4):403-406. [doi: 10.1001/jama.297.4.403] [Medline: 17244837]
29. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
30. Wellpass. URL: https://www.wellpass.com/company/ [accessed 2017-11-21] [WebCite Cache ID 6v91eWhNl]
31. Powell BJ, McMillen JC, Proctor EK, Carpenter CR, Griffey RT, Bunger AC, et al. A compilation of strategies for implementing clinical innovations in health and mental health. Med Care Res Rev 2012 Apr;69(2):123-157 [FREE Full text] [doi: 10.1177/1077558711430690] [Medline: 22203646]
32. Bureau of Labor Statistics. May 2016 National Occupational Employment Wage Estimates URL: https://www.bls.gov/oes/current/oes_nat.htm/ [WebCite Cache ID 6wY2LhLEq]
33. Healthcare Bluebook. URL: https://healthcarebluebook.com/ [accessed 2018-01-17] [WebCite Cache ID 6wY2Hmf0v]

34. Kolltveit BH, Gjengedal E, Graue M, Iversen MM, Thorne S, Kirkevold M. Conditions for success in introducing telemedicine in diabetes foot care: a qualitative inquiry. BMC Nurs 2017;16:2 [FREE Full text] [doi: 10.1186/s12912-017-0201-y] [Medline: 28100957]
35. Odnoletkova I, Buysse H, Nobels F, Goderis G, Aertgeerts B, Annemans L, et al. Patient and provider acceptance of telecoaching in type 2 diabetes: a mixed-method study embedded in a randomised clinical trial. BMC Med Inform Decis Mak 2016 Dec 09;16(1):142 [FREE Full text] [doi: 10.1186/s12911-016-0383-3] [Medline: 27825340]
36. Given JE, Bunting BP, O'Kane MJ, Dunne F, Coates VE. Tele-Mum: a feasibility study for a randomized controlled trial exploring the potential for telemedicine in the diabetes care of those with gestational diabetes. Diabetes Technol Ther 2015 Dec;17(12):880-888. [doi: 10.1089/dia.2015.0147] [Medline: 26394017]
37. Welch G, Balder A, Zagarins S. Telehealth program for type 2 diabetes: usability, satisfaction, and clinical usefulness in an urban community health center. Telemed J E Health 2015 May;21(5):395-403. [doi: 10.1089/tmj.2014.0069] [Medline: 25748544]
38. Li J, Talaei-Khoei A, Seale H, Ray P, Macintyre CR. Health care provider adoption of eHealth: systematic literature review. Interact J Med Res 2013;2(1):e7 [FREE Full text] [doi: 10.2196/ijmr.2468] [Medline: 23608679]
39. Gagnon M, Ngangue P, Payne-Gagnon J, Desmartis M. m-Health adoption by healthcare professionals: a systematic review. J Am Med Inform Assoc 2016 Jan;23(1):212-220. [doi: 10.1093/jamia/ocv052] [Medline: 26078410]

Abbreviations

BH: Bellevue Hospital
FBG: fasting blood glucose
GH: Gouverneur Health
HbA_{1c}: glycated hemoglobin
MITI: mobile insulin titration intervention
OID: optimal insulin dose
SMS: short message service

Edited by G Eysenbach; submitted 27.11.17; peer-reviewed by E Burner, L Nelson, A Menon; comments to author 21.12.17; revised version received 25.01.18; accepted 12.02.18; published 19.03.18.

Please cite as:
Levy NK, Orzeck-Byrnes NA, Aidasani SR, Moloney DN, Nguyen LH, Park A, Hu L, Langford AT, Wang B, Sevick MA, Rogers ES
Transition of a Text-Based Insulin Titration Program From a Randomized Controlled Trial Into Real-World Settings: Implementation Study
J Med Internet Res 2018;20(3):e93
URL: http://www.jmir.org/2018/3/e93/
doi:10.2196/jmir.9515
PMID:29555621

Original Paper

Using Fitness Trackers and Smartwatches to Measure Physical Activity in Research: Analysis of Consumer Wrist-Worn Wearables

André Henriksen[1], MSc (Comp Sci), MBA; Martin Haugen Mikalsen[2], BSc (Comp Sci); Ashenafi Zebene Woldaregay[2], MSc (Comp Eng), MSc (Telemed & e-Health); Miroslav Muzny[3,4], MSc (Comp Sci); Gunnar Hartvigsen[2], MSc (Comp Sci), PhD; Laila Arnesdatter Hopstock[5], RN, CRNA, MSc (Nursing), PhD; Sameline Grimsgaard[1], MPH, MD, PhD

[1]Department of Community Medicine, University of Tromsø – The Arctic University of Norway, Tromsø, Norway

[2]Department of Computer Science, University of Tromsø – The Arctic University of Norway, Tromsø, Norway

[3]Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway

[4]Spin-Off Company and Research Results Commercialization Center, 1st Faculty of Medicine, Charles University in Prague, Prague, Czech Republic

[5]Department of Health and Care Sciences, University of Tromsø – The Arctic University of Norway, Tromsø, Norway

Corresponding Author:
André Henriksen, MSc (Comp Sci), MBA
Department of Community Medicine
University of Tromsø – The Arctic University of Norway
Postboks 6050 Langnes
Tromsø, 9037
Norway
Phone: 47 77644000
Email: andre.henriksen@uit.no

Abstract

Background: New fitness trackers and smartwatches are released to the consumer market every year. These devices are equipped with different sensors, algorithms, and accompanying mobile apps. With recent advances in mobile sensor technology, privately collected physical activity data can be used as an addition to existing methods for health data collection in research. Furthermore, data collected from these devices have possible applications in patient diagnostics and treatment. With an increasing number of diverse brands, there is a need for an overview of device sensor support, as well as device applicability in research projects.

Objective: The objective of this study was to examine the availability of wrist-worn fitness wearables and analyze availability of relevant fitness sensors from 2011 to 2017. Furthermore, the study was designed to assess brand usage in research projects, compare common brands in terms of developer access to collected health data, and features to consider when deciding which brand to use in future research.

Methods: We searched for devices and brand names in six wearable device databases. For each brand, we identified additional devices on official brand websites. The search was limited to wrist-worn fitness wearables with accelerometers, for which we mapped brand, release year, and supported sensors relevant for fitness tracking. In addition, we conducted a Medical Literature Analysis and Retrieval System Online (MEDLINE) and ClinicalTrials search to determine brand usage in research projects. Finally, we investigated developer accessibility to the health data collected by identified brands.

Results: We identified 423 unique devices from 132 different brands. Forty-seven percent of brands released only one device. Introduction of new brands peaked in 2014, and the highest number of new devices was introduced in 2015. Sensor support increased every year, and in addition to the accelerometer, a photoplethysmograph, for estimating heart rate, was the most common sensor. Out of the brands currently available, the five most often used in research projects are Fitbit, Garmin, Misfit, Apple, and Polar. Fitbit is used in twice as many validation studies as any other brands and is registered in ClinicalTrials studies 10 times as often as other brands.

Conclusions: The wearable landscape is in constant change. New devices and brands are released every year, promising improved measurements and user experience. At the same time, other brands disappear from the consumer market for various reasons. Advances in device quality offer new opportunities for research. However, only a few well-established brands are frequently used in research projects, and even less are thoroughly validated.

(J Med Internet Res 2018;20(3):e110) doi:10.2196/jmir.9157

KEYWORDS

motor activity; physical activity; fitness trackers; heart rate; photoplethysmography

Introduction

Background

The World Health Organization recommends 150 min of moderate intensity physical activity (PA) each week for adults and 60 min for children and adolescents [1]. However, 25% of adults and more than 80% of adolescents do not achieve the recommended PA targets [1]. Results from the Tromsø Study, the longest running population study in Norway, shows that only 30.4% of women and 22.0% of men reach the recommended target [2].

Low PA is currently the fourth leading risk factor for mortality worldwide [3]. Even though there is limited evidence that using wearable fitness trackers will improve health [4,5], these devices are still popular, and new fitness devices appear on the consumer market regularly. In 2016, vendors shipped 102 million devices worldwide, compared with 82 million in 2015 [6]. Fifty-seven percent of these devices were sold by the top five brands: Fitbit, Xiaomi, Apple, Garmin, and Samsung. The first quarter of 2017 shows an increase of 18% in devices sold, compared with the same period in 2016 [7]. With a large number of available devices and brands, it is difficult to navigate through an ever-growing list of brands and devices with different capabilities, price, and quality.

Available sensors and internal interpreting algorithms determine device output. Sensor data are, in most devices, reduced to a limited set of metrics before being transferred to the user's mobile phone. In addition, limited space affects how long the device can collect data before such a transfer is needed. Data are stored locally, and in many cases, uploaded to brand specific or open cloud–based health repositories. Accessing these data by third-party apps and comparing them is not always possible. These interoperability challenges were recently identified in a study by Arriba-Pérez et al [8]. They suggested ways to handle these issues, but they did not make any brand or device recommendations. Several studies have compared activity-tracking wearables. As an example, Kaewkannate and Kim [9] did a comparison of four popular fitness trackers in 2016. They compared devices objectively and subjectively. Data were thoroughly collected, but because of the rapid release of new devices, these four devices will be among the most popular only for a relatively short time. A comparison of brands is also of interest because brands from larger companies are, compared with small start-ups and crowd funded brands, likely to survive longer. In addition, it is of interest to know which brands support the various available programming options. Sanders et al [10] did a literature review on articles using wearables for health self-monitoring and sedentary behavior and PA detection. They reviewed various aspects of these devices, but they gave no details about device sensor support and suitability in research.

The objective of this study was to examine how the consumer market for wearables has evolved, and analyze and summarize available devices that can measure PA and heart rate (HR). Moreover, we aim to identify brands that are used extensively in research projects, and compare and consider their relevance for future studies.

Sensors

A plethora of devices promises to measure PA in new and improved ways. These devices use different sensors and algorithms to calculate human readable metrics based on sensor output. Traditional step counters use pedometers to detect daily step counts. Although cheap and energy efficient, pedometers are not as accurate as accelerometers, which is the current standard for collecting PA data [11]. All modern fitness trackers and smartwatches have an accelerometer. Compared with research tools (eg, ActiGraph [12]), these devices are considered less accurate for some measurements [13,14]. However, they are generally less invasive, cheaper, have more functionality, are more user-friendly, and are increasingly being used in research. Most accelerometer-based fitness wearables measure acceleration in three directions [15] and can be used to estimate type of movement, count steps, calculate energy expenditure (EE) and energy intensity, as well as estimate sleep patterns and more. The validity and reliability of these metrics varies. Evenson et al [14] did a review in 2015 and found high validity for steps but low validity for EE and sleep. Furthermore, they found reliability for steps, distance, EE, and sleep to be high for some devices.

In addition, some wearables have gyroscopes, magnetometers, barometers, and altimeters. A gyroscope can potentially increase device accuracy by measuring gravitational acceleration, that is, orientation and angular velocity, and better estimate which activity type a person is performing [16]. A magnetometer is a digital compass [15] and can improve motion tracking accuracy by detecting the orientation of the device relative to magnetic north. Magnetometers improve accuracy by compensating for gyroscope drift, a problem with gyroscopes where the rotation axis slowly drifts from the actual motion and must be restored regularly. Accelerometers, gyroscopes, and magnetometers are often combined into an inertial measurement unit (IMU). Most mobile phones use IMUs to calculate orientation, and an increasing number of fitness wearables include this unit to give more accurate metrics. Barometers or altimeters detect changes in altitude [15] and can be used to improve some metrics (eg, EE), as well as report additional metrics (eg, climbed floors).

Photoplethysmography (PPG) is a relatively new technique in wearables. PPG is an optical technique to estimate HR by monitoring changes in blood volume beneath the skin [17]. A light-emitting diode projects light onto the skin, which is affected by the HR and reflected back to the sensor. However, movement, ambient light, and tissue compression affect the light, resulting in signal noise, and cleaning algorithms often

use accelerometer data to assist HR estimation [18]. There is some evidence that gyroscopes could be used [19] to reduce PPG signal noise, so we are likely to see more devices in the future equipped with PPG sensors. To further enrich the PA data collection, some devices have a built in global positioning system (GPS) receiver. This is especially true for high-end fitness trackers and sports watches specifically targeting physically active people. With a GPS, it is possible to track more data, including position, speed, and altitude.

Algorithms and Mobile Apps

Raw data from sensors must be converted into readable metrics to be meaningful for the user. Many devices only display a limited set of metrics directly on the device (eg, today's step count or current HR) and rely on an accompanying mobile app to show the full range of available metrics (eg, historic daily step count and detailed HR data). Although the physical sensors in these devices are very similar, the algorithms that interpret sensor output are unique for most vendors. These algorithms are often company secrets, and they can be changed without notice. In addition, the quality and supported features of the accompanying mobile apps varies, and the total user experience will therefore differ. Each additional sensor included in a device can be used to add additional types of metrics for the user or supply internal algorithms with additional data to improve accuracy of already available metric types. However, additional sensors affect price and power consumption.

Device Types

There are many similarities between different types of devices, and they may be difficult to categorize. We will use the term wearable in this paper as a common term for wrist-worn devices that can track and share PA data with a mobile phone.

A smartwatch is a wrist-worn device that, mostly, acts as an extension to a mobile phone and can show notifications and track PA and related metrics. Modern smartwatches often include a touch screen and can support advanced features and display high resolution activity trends [15]. Fitness trackers (ie, smart band or fitness band), normally worn on the wrist or hip, are devices more dedicated to PA tracking. A fitness tracker is typically cheaper than a smartwatch because of less expensive hardware and often fewer sensors. Due to this, it generally also has better battery life and a limited interface for displaying tracking results [15].

Other terms are also used, for example, sports watch and GPS watch, which can be considered merges between smartwatches and fitness trackers. In addition, there are hybrid watches (ie, hybrid smartwatches) that have a traditional clockwork and analogue display that have been fitted with an accelerometer. An accompanying mobile app is needed to access most data, but daily step counts are often represented as an analogue gauge on the watch face.

Wearable Usage Scenario

Wearables come forward as a new alternative to tracking PA in research (compared with, eg, ActiGraph), especially when it is desired to collect measurements for a prolonged period of time. In an intervention study, continuous data collecting from wearables would allow researchers to better track changes in PA and adjust the intervention accordingly. Wearables can also be used in epidemiological research as a tool for tracking PA for an extended period. This could reveal detailed PA changes in a population over time. In both scenarios, there are several potential important requirements to consider when choosing a device for the study, including usability, battery life, price, accuracy, durability, look and feel, and data access possibilities.

Methods

Search Strategies

Brands, Devices, and Sensors

We searched six databases to create a list of relevant wearable devices: The Queen's University's Wearable Device Inventory [20], The Vandrico Wearables database [21], GsmArena [22], Wearables.com [23], SpecBucket [24], and PrisGuide [25,26]. We only used publicly available information when comparing devices. We did the search from May 15, 2017 to July 1, 2017.

We identified wearables in two steps. In step one, we identified and searched the six defined databases. In step two, we extracted all brands from the list of devices identified in step one and examined brand websites for additional devices. If we found the same device in several databases with conflicting information, we manually identified the correct information from the device's official website or other online sources (eg, Wikipedia and Google search). We removed duplicates and devices not fitting the inclusion criteria.

Brand Usage in Research

We searched Ovid MEDLINE on September 30, 2017 to determine how often the most relevant brands were used in previous studies. For each search, we performed a keyword search with no limitations set. We divided our findings into validation and reliability studies and data collection studies.

To decide which brand to consider most relevant, we did two sets of searches. In the first set, we created a brand-specific keyword search for brands that were (1) One of the five most sold brands in 2015 or 2016 or (2) Had released 10 or more unique devices. From the resulting list of articles, we screened title, abstract, and the method section. This screening was done to (1) Exclude articles out of scope and (2) To identify additional brands used in these studies. We compiled a list of these brands and performed a second set of searches, one for each new identified brand. Eleven brands were finally included. The specific keyword search used for each brand is given in the Results section where we summarize our findings.

We also searched the US National Library of Medicine database of clinical studies through the ClinicalTrials website, using the same 11 keyword searches, to determine brand usage in ongoing projects. One author did the articles screening, as well as the projects description screening in ClinicalTrials.

Brand Developer Possibilities

To determine how relevant a specific brand is when planning a new research project, we reviewed the 11 identified brands and considered available developer options, supported mobile

phone environments, and options for health data storage. We especially reviewed availability of an application programming interface (API) and a software development kit (SDK). Information was collected from Google Play, Apple's App Store, and official brand websites. Information retrieval was done in September 2017.

Inclusion and Exclusion Criteria

Brands, Devices, and Sensors

The study is limited to wrist-worn consumer devices that utilize accelerometers to measure PA. Devices capable of collecting HR from the wrist using an optical sensor were tagged as PPG devices. Devices were tagged as GPS devices only if they had a built-in GPS tracker. We only included devices meant for personal use, designed to be worn continuously (24/7), and were capable of sharing data with mobile phones through Bluetooth. The wrist-worn limitation was added because hip-worn devices are not normally worn during the night (ie, not 24/7). Only devices released before July 1, 2017 were included. We excluded hybrid watches because most hybrid vendors make a large number of watch variations, with what seems to be the same hardware. In addition, these watches are mostly available through high-end suppliers of traditional watches, at a price point that would prevent researchers from considering their use in a large study.

Brand Usage in Research

Due to the large number of available brands, we limited our search to include only the 11 brands already identified as relevant. We excluded brands that are no longer available (ie, company shut down). Review studies were also excluded.

Brand Developer Possibilities

When reviewing brand relevance in research, we only reviewed developer capabilities for the 11 brands we had already included in the list of relevant brands. We set the additional limitation that the brand was used in at least one article in Ovid MEDLINE.

Device Categorization, Data Collection, and Reporting Categories

When collecting information about wearables, we categorized them into three groups:

1. Smartwatches: a device was tagged as a smartwatch if
   - It supported mobile phone notifications, and the vendor described it as a smart watch, or if
   - It had a touch screen and was not explicitly described as a fitness tracker by the vendor.
2. Fitness trackers: we classified a device as a fitness tracker if
   - Its main purpose was to track PA, or if
   - The vendor called it a fitness tracker, or if
   - The device did not support notifications from the connected mobile phone (eg, incoming calls or texts).
3. Hybrid watches: to be considered a hybrid watch, the device had to have an analogue clockwork with a built-in digital accelerometer.

We collected the following variables for each device: brand name, device name, year of release, country of origin, device type (eg, fitness tracker), and whether they had a built-in accelerometer, gyroscope, magnetometer, barometer or altimeter, GPS, and PPG.

We looked at three aspects of the devices we identified and reported under three categories:

1. Metrics and trends: in this category, we described the status for available brands, devices, and sensors, as well as reviewed trends in sensor availability over time.
2. Brand usage in research: in this category, we searched Ovid MEDLINE and ClinicalTrials and determined which brands are most used in a research setting.
3. Brand developer possibilities: in this category, we reviewed software integration platforms and mobile platform support for the most relevant brands.

Results

Relevant Devices

An overview of the device search process is given in Figure 1. We found 572 devices by searching online and offline databases and 131 additional devices by visiting the official websites for each identified brand, totaling 703 devices. Removing duplicates left 567 unique devices. These were screened for variation, that is, the same device with different design. After excluding 41 because of variation, 526 remained and were screened for eligibility. We removed 103 devices for not fitting the inclusion criteria. The remaining 423 devices were included in the study.

Brands, Devices, and Sensors

Brands

We identified 423 unique wearables, distributed between 132 different brands. Almost half the brands (47.0%, 62/132) had only one device. Moreover, 75.0% (99/132) of brands had three or fewer devices, and 83.3% (110/132) had five or fewer devices. Brands originated from 23 different countries, but the United States (43.2%, 57/132) and China (16.7%, 22/132, mainland China; 19.0%, 25/132, including Taiwan) represented the largest number of brand origin. Each remaining country represented between 0.8% (1/132) and 5.3% (7/132) of brands.

As the market has grown and wearable technology has become increasingly popular, a number of new brands have appeared on the market. In 2011, there were only three brands available. There was a small increase in brand count in 2012 and 2013, but in 2014, we saw the largest increase with 41 new brands. The number of new brands started to decrease in 2015, with 36 new brands in 2015 and 23 in 2016. Only three new brands have been introduced in 2017, but this number only represents the first 6 months of 2017. The final count for 2017 will likely be higher. An overview of the number of new brands that appeared on the market between 2011 and 2017 is given in Figure 2. Note that some companies are no longer active and, for 17 devices, we could not determine release year.

Most brands only had a small number of wearables, but some produced a lot more. The brand with most unique wearables

was Garmin (United States) with 40 different devices. No.1 (China) introduced the second highest number of wearables with 19 devices. An overview of the release year of the 22 (out of 132) brands that have released more than five devices is given in Table 1. Seven out of these 22 brands originated in the United States, five (six including Taiwan) originated in China, and two originated in South Korea. All other countries are represented only once. Some of these brands are no longer active (eg, Pebble and Jawbone).

Devices

Three devices were released in 2011 (earliest year), seven in 2012, 30 in 2013, and 87 in 2014. The year with the highest number of new wearables was 2015, with 121 new devices. In 2016, 120 new devices were released; the first year with a decreasing number of new wearables. The number of new and accumulated devices from 2011 to 2017 is summarized in Table 2. The last column (unknown) represents devices where we could not identify the release year. The above numbers represent the total number of new devices. If grouped into fitness trackers and smartwatches, there is a small overrepresentation among new smartwatches. Up until 2014, about half of devices were smartwatches. In 2015 and 2016, smartwatches represented 59.3% (143/241) of new devices, whereas fitness trackers represented 40.6% (98/241).

Sensors

The number of sensors included in new devices have increased in the last few years. Since 2015, the order of the most common sensors has consistently been PPG, GPS, gyroscope, magnetometer, and barometer or altimeter. In addition, these sensors have had a steady increase in availability in the same period. For 2017, 71% (27/38) of new devices included a PPG sensor, 50% (19/38) included a GPS, 39% (15/38) included a gyroscope, 34% (13/38) included a magnetometer, and 32% (12/38) included a barometer or altimeter. Figure 3 gives an overview of the number of devices each year that includes each sensor, in percent of total number of released devices that year. Devices with more than one sensor are represented once for each sensor it includes.

In total, since 2011, 38.5% (163/423) of wearables have only been equipped with one sensor (accelerometer). Moreover, 29.8% (126/423) of devices had two sensors, 12.1% (51/423) had three sensors, 11.1% (47/423) had four sensors, and 6.4% (27/423) had five sensors. Only 2.1% (9/423) of devices had all six sensors. In Table 3, these numbers are broken down by sensor combination and year. Some sensor combinations do not exist and are excluded.

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

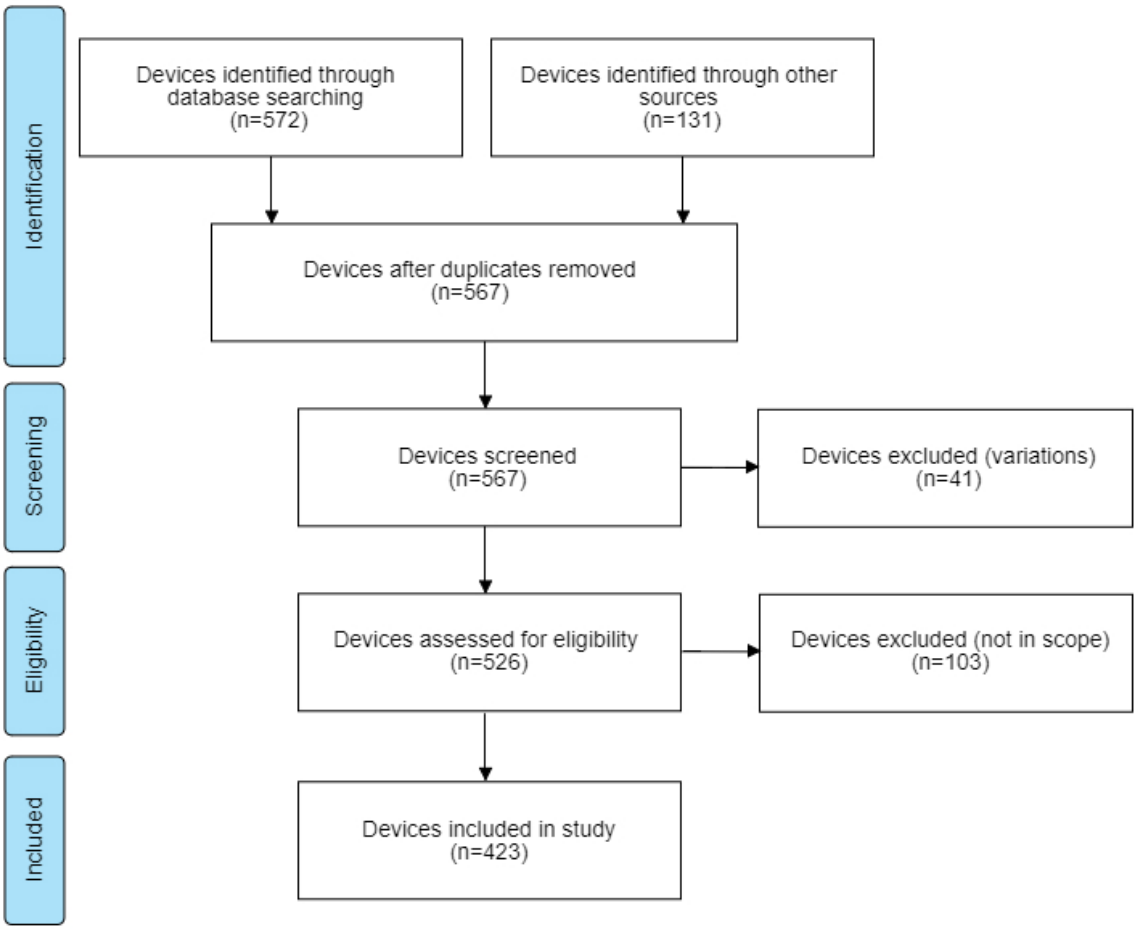


Figure 2. Number of new and aggregated available brands by year.

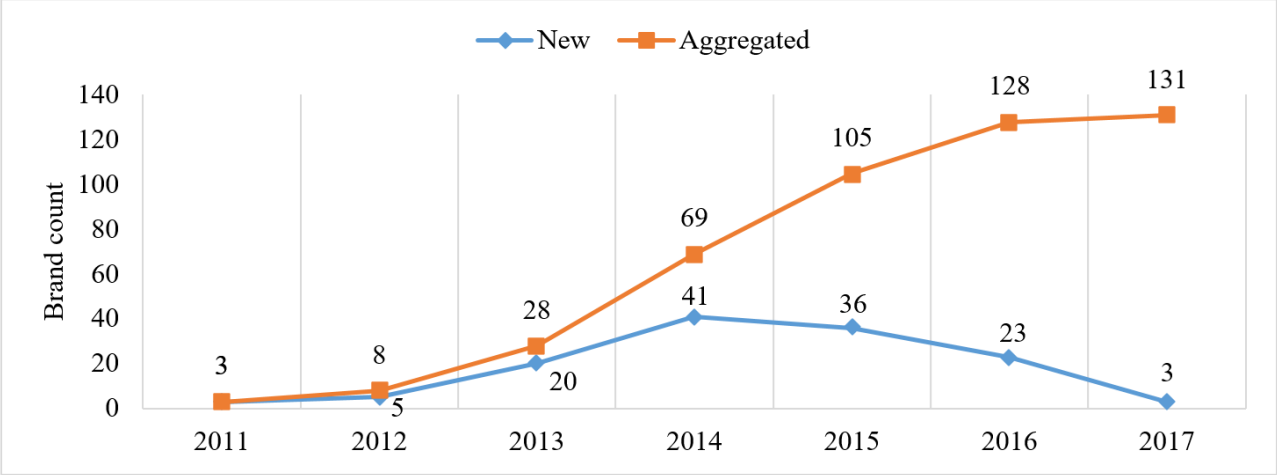


Table 1. Device count per year for brands with six or more wearables.

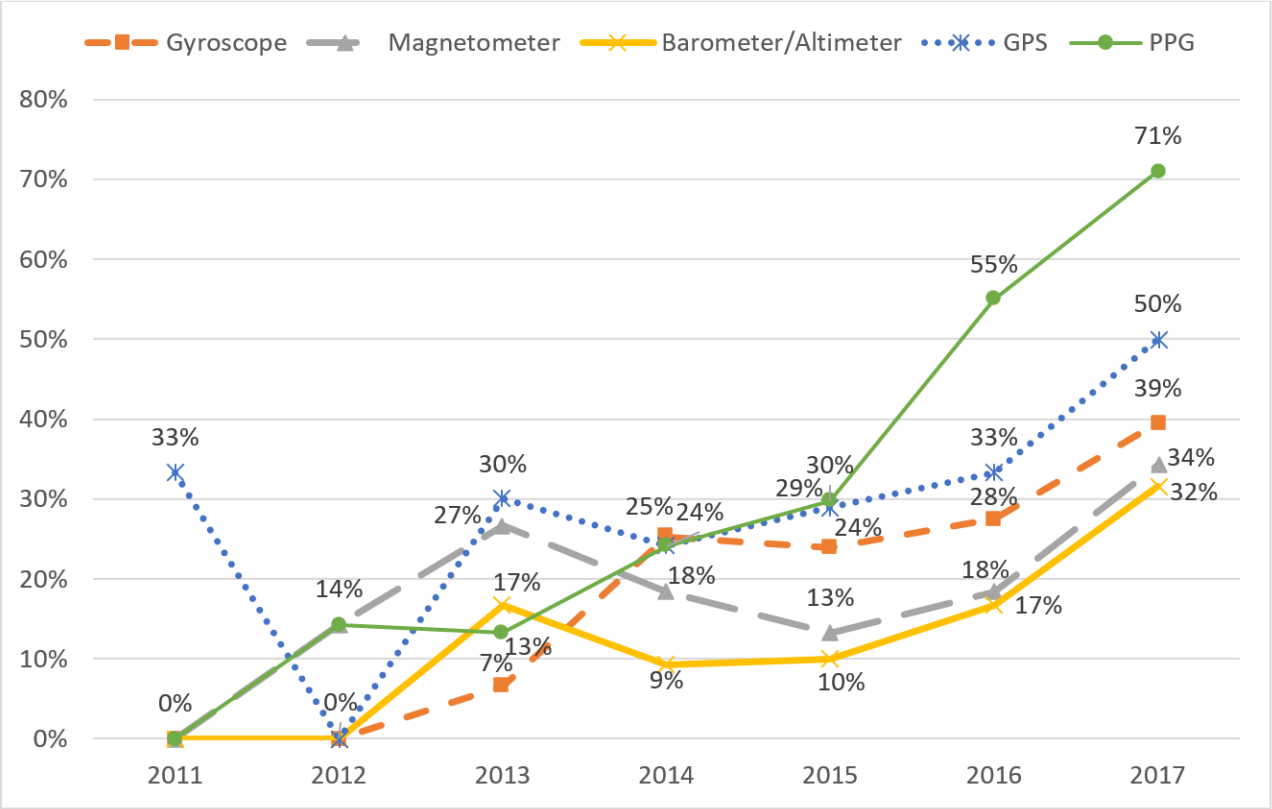
Brand	Country	2011	2012	2013	2014	2015	2016	2017	Unknown	Total[a]
Garmin	United States		1	5	6	11	13	4		40
Fitbit	United States			1	1	2	4	1		9
Misfit	United States			1	1	3	1	2		8
LifeTrak	United States			1	5		1			7
iFit	United States				1	4	1			6
Jawbone	United States	1		1	1	3				6
Pebble	United States			1	1	3	1			6
No. 1	China					5	9	5		19
Omate	China				2	5	2			9
Zeblaze	China					2	5	2		9
Huawei	China				1	3	3	1		8
Oumax	China				1	2	2	1	1	7
Mobile Action	Taiwan					2	2		4	8
Samsung	South Korea			1	6	1	4			12
LG	South Korea				3	1	1	2		7
WorldSim	England					1	1		5	7
Polar	Finland			1	2	4	2	2		11
Technaxx	Germany				4		2			6
Awatch	Italy						3	4		7
Epson	Japan				2	5				7
TomTom	Netherlands				2	1	4			7
MyKronoz	Switzerland				4	6	7	1		18

[a]Total brand count for the United States=7, China and Taiwan=6, and South Korea=2. All other countries are represented only once.

Table 2. Number of new and accumulated devices by year.

Devices	2011	2012	2013	2014	2015	2016	2017	Unknown
New	3	7	30	87	121	120	38	17
Accumulated	3	10	40	127	248	368	406	423

Figure 3. Percentage of devices released each year, supporting each sensor. GPS: global positioning system; PPG: photoplethysmography.



Brand Usage in Research

The top five vendors in 2015 [27] and 2016 [6], in sold units, were Fitbit, Xiaomi, Apple, Garmin, and Samsung. Brands with more than 10 unique wearables include Garmin, No.1, MyKronoz, Samsung, and Polar. These eight, and additional brands identified during the MEDLINE search and ClinicalTrials search, were considered. We did not find any publications or active clinical trials that used devices from No.1 or MyKronoz. Devices from Basis, BodyMedia, Pebble, Jawbone, Microsoft, and Nike were also used in some of the identified studies, but these brands do no longer produce wearables within the scope of this paper and were excluded from further analysis.

The MEDLINE search resulted in 81 included studies that we divided into two groups: (1) validation and reliability studies and (2) data collection studies. Studies where wearable output was compared with existing research instruments known to give accurate results (eg, ActiGraph) or with direct observation, as well as studies where several wearables were compared with each other for accuracy or reliability, were classified as validation and reliability studies. Studies where wearables were used as a tool for intervention or observation, to collect data on PA, HR, EE, sleep, or other available metrics, were classified as data collection studies. Out of these 81 studies, 61 were classified as validation and reliability studies, whereas 20 were classifies as data collection studies.

Fitbit devices were used in 54 studies [9,13,28-79]. Out of these, 40 studies were validation or reliability studies. In 22 of the studies, one or more Garmin devices were used [32,33,46,49,50,62,77-92]. Of these, 18 were validation or reliability studies. Eight studies used Apple devices [29,30,35,49,62,79,93,94]. Six of these were validation or reliability studies. All studies using devices from Misfit, Polar, Withings, Mio, Samsung, PulseOn, TomTom, and Xiaomi were validation or reliability studies. Misfit devices were used in 12 studies [9,36,42,43,46,61-63,85,95-97]; Polar devices were used in 6 studies [36,43,46,62,98,99]; Withings [63,85,89,100,101], Mio [29,30,54,102,103], and Samsung [29,30,58,62,96] devices were used in 5 studies; PulseOn devices were used in 4 studies [29,104-106]; TomTom devices were used in 2 studies [54,79]; and Xiaomi devices were used in 1 study [96].

From ClinicalTrials, we found that the vast majority of ongoing projects use, or are planning to use, Fitbit devices. All other devices were mentioned in three or less projects, whereas Fitbit devices were mentioned in 31 studies. A summary of these studies and projects is given in Table 4. We further grouped the validation and reliability studies into five categories. A total of 31 studies focused on step counts or distance, 15 studies researched EE, 15 studies measured HR, 10 studies measured sleep, and 7 studies collected other metrics. Multimedia Appendix 1 gives an overview of articles found in MEDLINE, which brands they included in the study, and which of the five categories they are grouped into.

Brand Developer Possibilities

Next, we considered developer possibilities for the 11 brands already identified as most relevant in research: Apple, Fitbit, Garmin, Mio, Misfit, Polar, PulseOn, Samsung, TomTom, Withings, and Xiaomi. All brands had an app in the Apple App Store and could connect to the iPhone. Except for the Apple Watch, all other brands had an app in Google Play and could be used with Android phones.

Table 3. Number and percentage of devices supporting a specific group of sensors, by year.

Sensors	2011	2012	2013	2014	2015	2016	2017
Accelerometer (Acc), n (%)	2 (67)	5 (71)	16 (53)	40 (46)	50 (41.3)	37 (30.8)	4 (11)
Acc + 1 sensor, n (%)							
PPG[a]		1 (14)	1 (3)	9 (10)	11 (9.1)	27 (22.5)	10 (26)
GPS[b]	1 (33)		2 (7)	9 (10)	15 (12.4)	3 (2.5)	
Gyroscope (Gyro)			1 (3)	3 (3)	9 (7.4)	4 (3.3)	1 (3)
Magnetometer (Mag)		1 (14)	2 (7)	1 (1)	3 (2.5)		
Barometer (Bar)				1 (1)	1 (0.8)		2 (5)
Acc + 2 sensors, n (%)							
GPS + PPG			1 (3)		7 (5.8)	6 (5)	3 (8)
Gyro + PPG				4 (5)	5 (4.1)	5 (4.2)	1 (3)
Gyro + GPS				1 (1)	2 (1.7)	2 (1.7)	
Bar + PPG			1 (3)		1 (0.8)	2 (1.7)	
Gyro + Mag				2 (2)	1 (0.8)		
Mag + GPS			1 (3)	1 (1)	1 (0.8)		
Mag + PPG						1 (0.8)	
Gyro + Bar				1 (1)			
Bar + GPS				2 (2)			
Acc + 3 sensors, n (%)							
Gyro + Mag + GPS			1 (3)	3 (3)	3 (2.5)	2 (1.7)	1 (3)
Gyro + Mag + PPG				4 (5)	2 (1.7)	3 (2.5)	1 (3)
Mag + Bar + GPS			3 (10)	2 (2)		4 (3.3)	1 (3)
Gyro + GPS + PPG				1 (1)		6 (5)	1 (3)
Bar + GPS + PPG						2 (1.7)	2 (5)
Mag + GPS + PPG						1 (0.8)	1 (3)
Gyro + Bar + PPG					2 (1.7)		
Gyro + Mag + Bar						1 (0.8)	
Acc + 4 sensors, n (%)							
Mag + Bar + GPS + PPG			1 (3)		3 (2.5)	4 (3.3)	
Gyro + Mag + GPS + PPG				1 (1)		3 (2.5)	3 (8)
Gyro + Bar + GPS + PPG					2 (1.7)	4 (3.3)	1 (3)
Gyro + Mag + Bar + GPS						1 (0.8)	2 (5)
Gyro + Mag + Bar + PPG				1 (1)	1 (0.8)		
Acc + 5 sensors, n (%)							
All sensors				1 (1)	2 (1.7)	2 (1.7)	4 (11)
Total, n	3	7	30	87	121	120	38

[a]PPG: photoplethysmography.

[b]GPS: global positioning system.

Table 4. Number of identified articles in Medical Literature Analysis and Retrieval System Online (MEDLINE) and ClinicalTrials.

Brand	MEDLINE[a] search term	MEDLINE		ClinicalTrials	
		Validation or reliability studies[b] (total article count=61)	Data collection studies[c] (total article count=20)	Validation or reliability studies[d]	Data collection studies[e]
Fitbit	Fitbit AND (Alta OR Blaze OR Charge OR Flex OR Surge)	40	14	1	30
Garmin	Garmin AND (Approach OR D2 OR Epix OR Fenix OR Forerunner OR Quatix OR Swim OR Tactix OR Vivo*)	18	4	1	2
Misfit	Misfit AND (Flare OR Flash OR Link OR Ray OR Shine OR Vapor)	12	0	0	1
Apple	Apple watch	6	2	1	1
Polar	Polar AND (“Polar Loop” OR M200 OR M4?0 OR M600 OR V800 OR A3?0)	6	0	1	3
Withings	Withings	5	0	0	2
Mio	Mio Alpha OR Mio Fuse OR Mio Slice	5	0	1	2
Samsung	Samsung Gear NOT “Gear VR” NOT Oculus	5	0	0	2
PulseOn	PulseOn	4	0	0	1
TomTom	TomTom	2	0	1	
Xiaomi	Xiaomi	1	0	0	1

[a]MEDLINE: Medical Literature Analysis and Retrieval System Online.

[b]Number of validation or reliability studies in MEDLINE.

[c]Number of data collection studies in MEDLINE.

[d]Number of validation or reliability studies in ClinicalTrials.

[e]Number of data collection studies in ClinicalTrials.

Three brands supported Windows Phone: Fitbit, Garmin, and Misfit. Apple Health and Google Fit are the two most common open cloud health repositories. Mio, Misfit, Polar, Withings, and Xiaomi, were the only brands that automatically synchronized fitness data to both of these repositories through these open APIs. The Apple Watch only synchronized automatically to the Apple Health repository. Seven out of 11 brands had a private cloud repository with an accompanying API, which allows third-party apps to access these data. Five brands had an SDK, which makes it possible to create custom programs to communicate with the device or create watch faces that can run on the device.

The Apple Watch was the only device running on watchOS. Three brands had at least one device running on Android Wear. The remaining seven brands used a custom system. A summary of all attributes for each brand is given in Table 5. Not all devices for a specific brand support all features. In addition, this is a snapshot of the status of these attributes, which are likely to change over time as new devices and brands expand their capabilities. The Apple Watch development environment is called WatchKit SDK and can be used to write apps for the Apple Watch [107]. Apple’s health storage solution is called Apple Health. A variety of different data types can be stored here and accessed by third-party developers through the HealthKit API [108]. Access to any of these services requires enrollment in the Apple Developer Program, which currently costs US $99 per year.

Fitbit offers three major SDKs (Device API, Companion API, and Settings API) for developing apps for Fitbit devices. In addition, Fitbit offers the Web API that can be used to access Fitbit cloud-stored fitness data. The Web API exposes six types of data: PA, HR, location, nutrition, sleep, and weight [109]. Fitbit also has a solution for accessing high-resolution step and HR data (ie, intraday data), granted on a case by case basis. There is no cost for developing with the Fitbit SDKs or API.

There are two generations of programmable Garmin wearables [110]. The Connect IQ SDK can be used by both generations, but devices using the newer Connect IQ 2 generation support more features. Development with this SDK is free. Garmin also offers a cloud-based Web API, Garmin Connect, which allows third-party apps to access users’ cloud-based fitness data. Access to this API costs US $5000 (one-time license). In addition, Garmin maintains a separate Health API intended to be used by companies for wellness improvement of their employees. This API is free but requires a manual approval from Garmin.

Table 5. Brand environment, integration, and development support.

Feature	Apple	Fitbit	Garmin	Mio	Misfit	Polar	PulseOn	Samsung	TomTom	Withings	Xiaomi
Supported platform											
Android		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
iPhone	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Windows phone		✓	✓		✓						
Integration											
Automatic synchronization to Apple Health	✓			✓	✓	✓				✓	✓
Automatic synchronization to Google Fit				✓	✓	✓				✓	✓
Private cloud storage		✓	✓		✓	✓		✓	✓	✓	
Cloud storage API[a]	✓	✓	✓		✓	✓		✓	✓	✓	
Developer SDK[b]	✓	✓	✓		✓			✓			
Watch system											
Android Wear					✓	✓					✓
watchOS (Apple)	✓										
Custom		✓	✓	✓			✓	✓	✓	✓	

[a]API: application programming interface.

[b]SDK: software development kit.

The Misfit developer ecosystem consists of three SDKs (Sleep SDK, Link SDK, and Device SDK) [111]. The Misfit Device SDK is the major SDK for developing apps for and communication with Misfit devices. This SDK is only available on request. Misfit also offers the Misfit Scientific Library that can be used to access Misfits proprietary sensor algorithms directly. This library is also only available on request. In addition, the Misfit Cloud API is used to access users' data from the Misfit cloud server. All SDKs and the API are free.

Polar does not offer a separate SDK. Polar devices can integrate with Google Fit and Apple Health and deposits collected data there [112]. This data are accessed using Google Fit APIs and Apple HealthKit APIs. In addition, data are uploaded to Polar's cloud storage, which is accessible by third-party developers through the AccessLink API. Besides PA data (steps, EE, and sleep), basic training data are also stored here. Access to AccessLink is free.

Development for a Samsung smartwatch is done using the Tizen SDK (Samsung smartwatch operating system is called Tizen). The Samsung Health SDK platform consists of two parts: Data SDK and Service SDK. Together these can be used to store and access health data collected from internal and external sensors, as well as third-party apps running on a Samsung watch or a mobile phone. Development using any of these services is free [113].

TomTom offers the Sports Cloud API for accessing data collected from TomTom devices. The API provides four types of data: PA (eg, exercises bouts), HR, tracking (eg, steps and EE), and physiology (eg, weight). Access to the API is free [114].

Nokia acquired Withings in 2016, and the original Withings API is now available as the Nokia Health API. Besides PA and sleep measurements, the API also gives access to intraday PA data. Nokia must manually approve access to this high-resolution activity API. The API is free [115].

Summarizing Results

Which features are most important when considering devices for a research project will depend on the purpose and design of the study. It is therefore not possible to identify one brand as the best brand in all circumstances. However, we have tried to quantify various aspects of a brand to identify and summarize their benefits.

Table 6. Brand summary.

Brand	Fitbit	Garmin	Misfit	Apple	Polar	Samsung	Withings	Mio	PulseOn	TomTom	Xiaomi	MyKronoz	No. 1
Devices[a]	9	40	8	3	11	12	2	3	1	7	3	18	19
MEDLINE[b]	54	22	12	8	6	5	5	5	4	2	1		
Validation or reliability[c]	40	18	12	6	6	5	5	5	4	2	1		
Steps	21	10	6	1	2	2	4				1		
Energy expenditure	10	4	3	4	3	1		2	2				
Heart rate	7	4	1	4	1	2		5	4	2	1		
Sleep	8	1	4		1		2						
Other	3	4	2		1								
ClinicalTrials[d]	31	3	1	2	4	2	2	3	1	1	1		
SDK[e]	✓	✓	✓	✓	✓	✓							
API[f]	✓	✓	✓	✓	✓	✓	✓			✓			
Apple Health[g]			✓	✓	✓		✓	✓			✓		
Google Fit[h]			✓		✓		✓	✓			✓		

[a]Number of unique devices.

[b]MEDLINE: Medical Literature Analysis and Retrieval System Online. Number of articles in MEDLINE.

[c]Number of validation or reliability studies in MEDLINE, grouped by metric (step, EE, HR, sleep, and others).

[d]Number of active projects in ClinicalTrials.

[e]Supports an SDK for third-party software implementation.

[f]API: application programming interface. Supports an API for developer access to data cloud.

[g]Supports automatic synchronization to Apple Health data cloud.

[h]Supports automatic synchronization to Google Fit data cloud.

We used eight categories in this custom comparison, which we suggest to consider before deciding on a brand for any research project:

1. Device count: a higher number of available devices make it possible to pick a device that is more tailored to the study.
2. Article count: a higher number of articles in Ovid MEDLINE indicate usage in previous studies.
3. Validation or reliability count: a high number of validation or reliability studies provides knowledge about device and brand accuracy.
4. ClinicalTrials count: a high number of active projects in ClinicalTrials indicate brand relevance.
5. SDK support: brands that allows third-party programs to run on their devices or communicate directly with the device, by offering an SDK, adds more possibilities for customization.
6. API support: brands that allows third-party programs to access the data cloud repository, by offering API access, adds more possibilities for health data collection and retrieval.
7. Apple Health: brands supporting automatic synchronization to Apple Health allow usage of Apple HealthKit API.
8. Google Fit: brands supporting automatic synchronization to Google Fit allow usage of Google Fit API.

A consensus between authors was reached to include these specific categories because we think together they indicate how often a specific brand has been used in the past and will be used in the future, and they show which options are available for data extraction. These are not the only possible categories, and each category will not be equally important for all studies.

Table 6 gives a summary of these categories for each brand. A transposed Excel (Microsoft) version for dynamic sorting is given in Multimedia Appendix 2. We have divided MEDLINE validation and reliability studies into subgroups, making it easier to compare brands for specific study purposes.

Discussion

Availability and Trends

The number of new brands increased every year from 2011 to 2014, but from 2015 to 2016, we saw a decrease in the number of new brands. The number of new devices also increased from 2011 to 2015, with a slight reduction in 2016. Many new and existing companies have tried to enter the wearable market during these years. Some have become popular, whereas others are no longer available. The number of new devices in the first two quarters of 2017 seems low, and there is a small indication that the number of new brands and devices released each year is declining. During the data collection phase, we also identified

a large number of hybrid watches. Although we did not report on these, this relatively new branch of wearables has grown in popularity. The Fossil group, representing 19 brands, recently announced they would launch more than 300 hybrid watches and smartwatches in 2017 [116]. Most of these will be hybrids, and 2017 may see the highest number of new hybrids released to date.

We only found nine devices that support all five sensors considered in this study. Among the 11 most relevant brands, only Fitbit Surge, Garmin Forerunner 935, Garmin Quatix 5, Samsung Gear S, and TomTom Adventure fall in this category. Most devices (68%) support only one sensor, in addition to the accelerometer. These numbers indicate that sensor count is not the main argument when choosing a device for personal use. In addition to the accelerometer, the most common sensors are PPG and GPS, regardless of sensor count. One reason for this may be that the added benefit of having these sensors, in a fitness setting, is very clear. Accelerometers can be used for step counting, PA intensity, exercise detection, and other well-understood metrics, whereas the added benefit of a gyroscope may be less intuitive. The added convenience of using a PPG compared with a pulse chest strap, or no HR detection at all, is also easy to understand. Adding a GPS also adds some easy-to-understand benefits, where tracking progress on a map and the possibility to detect speed is the most obvious. Magnetometers and barometers or altimeters may not be sensors that most people consider relevant for PA, although they can be used to enhance accuracy of EE and other metrics.

Brand Usage in Research

In the MEDLINE literature search, we found 81 studies that used one or more of the 11 brands we identified as most relevant in research. Out of these, 61 were validation or reliability studies. The remaining 20 studies used wearable devices as data collection instruments to measure PA, HR, EE, sleep, or other metrics. Fitbit was used in twice as many validation or reliability studies as any other brand. This has likely contributed to the high number of studies where Fitbit was used as the only instrument for health data collection. The same trend will likely continue in future publications because numbers from ClinicalTrials for active projects shows an overrepresentation of Fitbit-enabled projects. Of the brands currently available, the five most often used in research projects are Fitbit, Garmin, Misfit, Apple, and Polar. In addition, these brands have all existed for several years and have either released a large number of unique devices or shipped a large number of total devices. As such, they are likely to stay on the market for the near future.

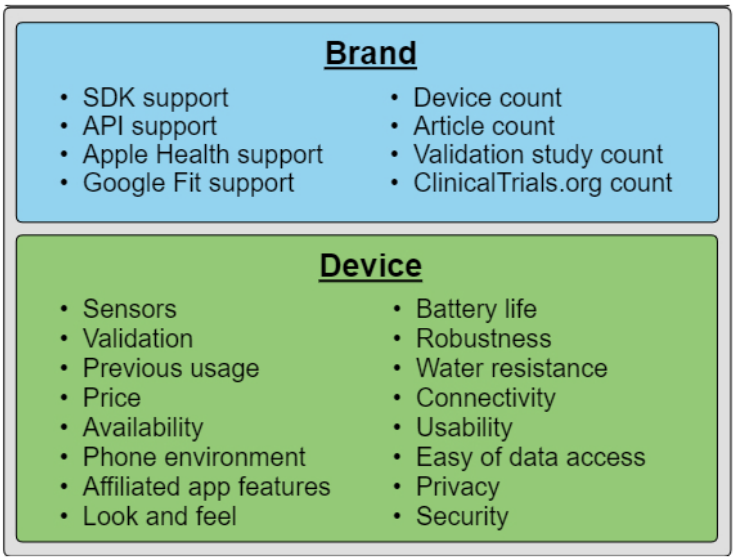
A high article count, high number of validation or reliability studies, or high number of studies in ClinicalTrials for a specific brand does not automatically imply validity or reliability. It does, however, show researcher interest in these brands.

Implication for Practice

Table 6 is a good starting point when considering brands for a new research project. Article count, validation or reliability study count, and ClinicalTrials count together indicate brand dependability. Larger numbers indicate how relevant, usable, and valid previous researchers have found each brand to be. In projects where it is relevant, SDK support allows programmatic interaction directly with the device. API support allows storage in, and access to, a brand-specific cloud-based health data repository. Apple Health and Google Fit support are alternative solutions for storing and accessing health data in an open cloud repository. For projects that require multiple brand support, using open solutions reduces the need to implement specific software for each brand. SDK, API, Apple Health, and Google Fit must be supported on both the brand and device level, however.

A high brand device count makes it easier to find a device that best supports the study needs. In addition to available sensors (ie, metrics), validation, and previous usage in research, several other potential relevant criteria exist, including price, availability, phone environment support, affiliated app features, look and feel, battery life, build quality or robustness, water resistance, connectivity, and usability.

Figure 4. Criteria to consider when choosing brand or device. API: application programming interface; SDK: software development kit.



Furthermore, projects that need programmatic access to the wearable or stored health data should especially consider SDK or API features and ease of use, as well as privacy and security. Figure 4 gives a summary of criteria to consider when selecting brand and device.

Limitations

We visited all the brands' websites to find additional devices, but several sites did not contain any information about discontinued devices. The release year of a device was rarely available on device webpages, and we had to search for reviews and other sources to find this information. The level of detail in device hardware specifications varied. Some vendors did not specify which sensor they included in their devices and only mentioned which features the device had. In some cases, the sensor could be derived from this information, but in other cases, we had to find this information elsewhere. Wikipedia was also used to collect sensor support and release year for some devices. This open editable encyclopedia is not necessarily always updated with correct information. For these reasons, there may be some inaccuracies in reported sensor support and release year. We did not collect information about device discontinuation. Reported numbers for total available devices does, therefore, not reflect the numbers of devices that currently can be store bought but rather the number of unique devices that have existed at some point.

Conclusions

In the last few years, we have seen a large increase in available brands and wearable devices, and more devices are released with additional sensors. However, for activity tracking, some sensors are more relevant than others are. In this study, we have focused on sensor support, health data cloud integration, and developer possibilities; because we find these to be most relevant for collection of PA data in research. However, deciding which wearable to use will depend on several additional factors.

The wearable landscape is constantly changing as new devices are released and as new vendors enter or leave the market, or are acquired by larger vendors. What currently are considered relevant devices and brands will therefore change over time, and each research project should carefully consider which brand and device to use. As a tool for future research, we have defined a checklist of elements to consider when making this decision.

Acknowledgments

The authors would like to thank Vandrico Solutions Inc for giving them API access to their database. They would also like to thank Steven Richardson and Debra Mackinnon at Queen's University for giving them an excerpt from their offline wearable database. The publication charges for this study have been funded by a grant from the publication fund of the University of Tromsø – The Arctic University of Norway.

Conflicts of Interest

None declared.

Multimedia Appendix 1

List of MEDLINE articles included in the results for "Brand usage in research".

[XLSX File (Microsoft Excel File), 24KB - jmir_v20i3e110_app1.xlsx]

Multimedia Appendix 2

Summary of the most important categories to consider when selecting a wearable brand for research.

[XLSX File (Microsoft Excel File), 13KB - jmir_v20i3e110_app2.xlsx]

References

1. World Health Organization. Physical activity URL: http://www.who.int/mediacentre/factsheets/fs385/en [accessed 2017-09-11] [WebCite Cache ID 6tOYSnXaX]
2. Emaus A, Degerstrøm J, Wilsgaard T, Hansen BH, Dieli-Conwright CM, Furberg AS, et al. Does a variation in self-reported physical activity reflect variation in objectively measured physical activity, resting heart rate, and physical fitness? Results from the Tromso study. Scand J Public Health 2010 Nov;38(5 Suppl):105-118. [doi: 10.1177/1403494810378919] [Medline: 21062845]
3. World Health Organization. 2017. Global Strategy on Diet, Physical Activity and Health URL: http://www.who.int/dietphysicalactivity/pa/en [accessed 2017-09-11] [WebCite Cache ID 6tOXoFxbF]
4. Finkelstein EA, Haaland BA, Bilger M, Sahasranaman A, Sloan RA, Nang EE, et al. Effectiveness of activity trackers with and without incentives to increase physical activity (TRIPPA): a randomised controlled trial. Lancet Diabetes Endocrinol 2016 Dec;4(12):983-995. [doi: 10.1016/S2213-8587(16)30284-4] [Medline: 27717766]
5. 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. J Am Med Assoc 2016 Sep 20;316(11):1161-1171. [doi: 10.1001/jama.2016.12858] [Medline: 27654602]

6. IDC. Wearables Aren't Dead, They're Just Shifting Focus as the Market Grows 16.9% in the Fourth Quarter, According to IDC URL: https://www.idc.com/getdoc.jsp?containerId=prUS42342317 [accessed 2018-03-01] [WebCite Cache ID 6xZyWGaxT]
7. IDC. Xiaomi and Apple Tie for the Top Position as the Wearables Market Swells 17.9% During the First Quarter URL: https://www.idc.com/getdoc.jsp?containerId=prUS42707517 [accessed 2018-03-01] [WebCite Cache ID 6xZypbatQ]
8. de Arriba-Perez F, Caeiro-Rodríguez M, Santos-Gago JM. Collection and processing of data from wrist wearable devices in heterogeneous and multiple-user scenarios. Sensors (Basel) 2016 Sep 21;16(9) [FREE Full text] [doi: 10.3390/s16091538] [Medline: 27657081]
9. Kaewkannate K, Kim S. A comparison of wearable fitness devices. BMC Public Health 2016 May 24;16:433 [FREE Full text] [doi: 10.1186/s12889-016-3059-0] [Medline: 27220855]
10. Sanders JP, Loveday A, Pearson N, Edwardson C, Yates T, Biddle SJ, et al. Devices for self-monitoring sedentary time or physical activity: a scoping review. J Med Internet Res 2016 May 04;18(5):e90 [FREE Full text] [doi: 10.2196/jmir.5373] [Medline: 27145905]
11. Corder K, Brage S, Ekelund U. Accelerometers and pedometers: methodology and clinical application. Curr Opin Clin Nutr Metab Care 2007 Sep;10(5):597-603. [doi: 10.1097/MCO.0b013e328285d883] [Medline: 17693743]
12. ActiGraph Corp. ActiGraph URL: http://actigraphcorp.com/ [accessed 2017-09-11] [WebCite Cache ID 6tOYrA9Ke]
13. Reid RE, Insogna JA, Carver TE, Comptour AM, Bewski NA, Sciortino C, et al. Validity and reliability of Fitbit activity monitors compared to ActiGraph GT3X+ with female adults in a free-living environment. J Sci Med Sport 2017 Jun;20(6):578-582. [doi: 10.1016/j.jsams.2016.10.015] [Medline: 27887786]
14. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015 Dec 18;12:159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
15. Richardson S, Mackinnon D. Left to their own devices? Privacy implications of wearable technology in Canadian workplaces URL: http://www.sscqueens.org/publications/left-to-their-own-devices [accessed 2018-03-01] [WebCite Cache ID 6xZzwLAN6]
16. Wagenaar RC, Sapir I, Zhang Y, Markovic S, Vaina LM, Little TD. Continuous monitoring of functional activities using wearable, wireless gyroscope and accelerometer technology. Conf Proc IEEE Eng Med Biol Soc 2011;2011:4844-4847. [doi: 10.1109/IEMBS.2011.6091200] [Medline: 22255423]
17. Allen J. Photoplethysmography and its application in clinical physiological measurement. Physiol Meas 2007 Mar;28(3):R1-39. [doi: 10.1088/0967-3334/28/3/R01] [Medline: 17322588]
18. Kim SH, Ryoo DW, Bae C. Adaptive noise cancellation Using accelerometers for the PPG signal from forehead. Conf Proc IEEE Eng Med Biol Soc 2007;2007:2564-2567. [doi: 10.1109/IEMBS.2007.4352852] [Medline: 18002518]
19. Casson AJ, Vazquez Galvez A, Jarchi D. Gyroscope vs. accelerometer measurements of motion from wrist PPG during physical exercise. ICT Express 2016 Dec;2(4):175-179. [doi: 10.1016/j.icte.2016.11.003]
20. Richardson SM, Mackinnon D. Queen's University. 2017. Wearable Device Inventory, Queen's University.
21. Vandrico. 2017. The wearables database URL: http://vandrico.com/wearables/ [accessed 2017-07-08] [WebCite Cache ID 6roNvEjdE]
22. GSM Arena. 2017. URL: http://www.gsmarena.com/results.php3?sFormFactors=8 [accessed 2017-07-08] [WebCite Cache ID 6roO1WCdv]
23. Wearables. 2017. Wearables.Com: Helping you make sense of wearable tech URL: http://www.wearables.com/devices/ [accessed 2017-07-08] [WebCite Cache ID 6roO94rOu]
24. Smartwatches. 2017. SpecBucket: Smart Watches URL: https://smartwatches.specbucket.com/ [accessed 2017-07-08] [WebCite Cache ID 6roOBKrxr]
25. Prisguiden. 2017. Smartklokke URL: https://prisguiden.no/kategorier/smartklokke [accessed 2017-07-08] [WebCite Cache ID 6roOHtXzf]
26. Prisguiden. 2017. Aktivitetsarmband URL: https://prisguiden.no/kategorier/aktivitetsarmband [accessed 2017-07-08] [WebCite Cache ID 6roOJt6HJ]
27. Business Wire. 2016. The Worldwide Wearables Market Leaps 126.9% in the Fourth Quarter and 171.6% in 2015 URL: https://www.businesswire.com/news/home/20160223005496/en/Worldwide-Wearables-Market-Leaps-126.9-Fourth-Quarter [accessed 2008-03-01] [WebCite Cache ID 6xa0wRrvV]
28. Balto JM, Kinnett-Hopkins DL, Motl RW. Accuracy and precision of smartphone applications and commercially available motion sensors in multiple sclerosis. Mult Scler J Exp Transl Clin 2016;2:2055217316634754 [FREE Full text] [doi: 10.1177/2055217316634754] [Medline: 28607720]
29. Shcherbina A, Mattsson CM, Waggott D, Salisbury H, Christle JW, Hastie T, et al. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. J Pers Med 2017 May 24;7(2) [FREE Full text] [doi: 10.3390/jpm7020003] [Medline: 28538708]
30. Wallen MP, Gomersall SR, Keating SE, Wisløff U, Coombes JS. Accuracy of heart rate watches: implications for weight management. PLoS One 2016;11(5):e0154420 [FREE Full text] [doi: 10.1371/journal.pone.0154420] [Medline: 27232714]

31. Chow JJ, Thom JM, Wewege MA, Ward RE, Parmenter BJ. Accuracy of step count measured by physical activity monitors: The effect of gait speed and anatomical placement site. Gait Posture 2017 Sep;57:199-203. [doi: 10.1016/j.gaitpost.2017.06.012] [Medline: 28666177]
32. Chen MD, Kuo CC, Pellegrini CA, Hsu MJ. Accuracy of wristband activity monitors during ambulation and activities. Med Sci Sports Exerc 2016 Dec;48(10):1942-1949. [doi: 10.1249/MSS.0000000000000984] [Medline: 27183123]
33. Duncan M, Murawski B, Short CE, Rebar AL, Schoeppe S, Alley S, et al. Activity trackers implement different behavior change techniques for activity, sleep, and sedentary behaviors. Interact J Med Res 2017 Aug 14;6(2):e13 [FREE Full text] [doi: 10.2196/ijmr.6685] [Medline: 28807889]
34. Sprint G, Cook D, Weeks D, Dahmen J, La Fleur A. Analyzing sensor-based time series data to track changes in physical activity during inpatient rehabilitation. Sensors (Basel) 2017 Sep 27;17(10) [FREE Full text] [doi: 10.3390/s17102219] [Medline: 28953257]
35. Chowdhury EA, Western MJ, Nightingale TE, Peacock OJ, Thompson D. Assessment of laboratory and daily energy expenditure estimates from consumer multi-sensor physical activity monitors. PLoS One 2017;12(2):e0171720 [FREE Full text] [doi: 10.1371/journal.pone.0171720] [Medline: 28234979]
36. Mercer K, Li M, Giangregorio L, Burns C, Grindrod K. Behavior change techniques present in wearable activity trackers: a critical analysis. JMIR Mhealth Uhealth 2016;4(2):e40 [FREE Full text] [doi: 10.2196/mhealth.4461] [Medline: 27122452]
37. Gaudet J, Gallant F, Bélanger M. A bit of fit: minimalist intervention in adolescents based on a physical activity tracker. JMIR Mhealth Uhealth 2017 Jul 06;5(7):e92 [FREE Full text] [doi: 10.2196/mhealth.7647] [Medline: 28684384]
38. Jacobsen RM, Ginde S, Mussatto K, Neubauer J, Earing M, Danduran M. Can a home-based cardiac physical activity program improve the physical function quality of life in children with fontan circulation? Congenit Heart Dis 2016;11(2):175-182. [doi: 10.1111/chd.12330] [Medline: 26879633]
39. Ridgers ND, Timperio A, Brown H, Ball K, Macfarlane S, Lai SK, et al. A cluster-randomised controlled trial to promote physical activity in adolescents: the Raising Awareness of Physical Activity (RAW-PA) Study. BMC Public Health 2017 Jan 04;17(1):6 [FREE Full text] [doi: 10.1186/s12889-016-3945-5] [Medline: 28052773]
40. Li LC, Sayre EC, Xie H, Clayton C, Feehan LM. A community-based physical activity counselling program for people with knee osteoarthritis: feasibility and preliminary efficacy of the track-OA study. JMIR Mhealth Uhealth 2017 Jun 26;5(6):e86 [FREE Full text] [doi: 10.2196/mhealth.7863] [Medline: 28652228]
41. Dondzila C, Garner D. Comparative accuracy of fitness tracking modalities in quantifying energy expenditure. J Med Eng Technol 2016 Aug;40(6):325-329. [doi: 10.1080/03091902.2016.1197978] [Medline: 27280592]
42. Bellone GJ, Plano SA, Cardinali DP, Chada DP, Vigo DE, Golombek DA. Comparative analysis of actigraphy performance in healthy young subjects. Sleep Sci 2016;9(4):272-279 [FREE Full text] [doi: 10.1016/j.slsci.2016.05.004] [Medline: 28154740]
43. Bai Y, Welk GJ, Nam YH, Lee JA, Lee JM, Kim Y, et al. Comparison of consumer and research monitors under semistructured settings. Med Sci Sports Exerc 2016;48(1):151-158. [doi: 10.1249/MSS.0000000000000727] [Medline: 26154336]
44. Lee HA, Lee HJ, Moon JH, Lee T, Kim MG, In H, et al. Comparison of wearable activity tracker with actigraphy for sleep evaluation and circadian rest-activity rhythm measurement in healthy young adults. Psychiatry Investig 2017 Mar;14(2):179-185 [FREE Full text] [doi: 10.4306/pi.2017.14.2.179] [Medline: 28326116]
45. Chu AH, Ng SH, Paknezhad M, Gauterin A, Koh D, Brown MS, et al. Comparison of wrist-worn Fitbit Flex and waist-worn ActiGraph for measuring steps in free-living adults. PLoS One 2017;12(2):e0172535 [FREE Full text] [doi: 10.1371/journal.pone.0172535] [Medline: 28234953]
46. Brooke SM, An HS, Kang SK, Noble JM, Berg KE, Lee JM. Concurrent validity of wearable activity trackers under free-living conditions. J Strength Cond Res 2017 Apr;31(4):1097-1106. [doi: 10.1519/JSC.0000000000001571] [Medline: 27465631]
47. Block VJ, Lizée A, Crabtree-Hartman E, Bevan CJ, Graves JS, Bove R, et al. Continuous daily assessment of multiple sclerosis disability using remote step count monitoring. J Neurol 2017 Feb;264(2):316-326. [doi: 10.1007/s00415-016-8334-6] [Medline: 27896433]
48. Morhardt DR, Luckenbaugh A, Goldstein C, Faerber GJ. Determining resident sleep during and after call with commercial sleep monitoring devices. Urology 2017 Aug;106:39-44. [doi: 10.1016/j.urology.2017.03.059] [Medline: 28502597]
49. Dooley EE, Golaszewski NM, Bartholomew JB. Estimating accuracy at exercise intensities: a comparative study of self-monitoring heart rate and physical activity wearable devices. JMIR Mhealth Uhealth 2017 Mar 16;5(3):e34 [FREE Full text] [doi: 10.2196/mhealth.7043] [Medline: 28302596]
50. Leth S, Hansen J, Nielsen OW, Dinesen B. Evaluation of commercial self-monitoring devices for clinical purposes: results from the Future Patient Trial, Phase I. Sensors (Basel, Switzerland) 2017;17(1):1-11.
51. Bian J, Guo Y, Xie M, Parish AE, Wardlaw I, Brown R, et al. Exploring the association between self-reported asthma impact and Fitbit-derived sleep quality and physical activity measures in adolescents. JMIR Mhealth Uhealth 2017 Jul 25;5(7):e105 [FREE Full text] [doi: 10.2196/mhealth.7346] [Medline: 28743679]
52. Coughlin SS, Hatzigeorgiou C, Anglin J, Xie D, Besenyi GM, De Leo G, et al. Healthy lifestyle intervention for adult clinic patients with type 2 diabetes mellitus. Diabetes Manag (Lond) 2017;7(2):197-204 [FREE Full text] [Medline: 28794802]

53. Åkerberg A, Koshmak G, Johansson A, Lindén M. Heart rate measurement as a tool to quantify sedentary behavior. Stud Health Technol Inform 2015;211:105-110. [Medline: 25980855]
54. Stahl SE, An HS, Dinkel DM, Noble JM, Lee JM. How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough? BMJ Open Sport Exerc Med 2016;2(1):1-7 [FREE Full text] [doi: 10.1136/bmjsem-2015-000106]
55. Alinia P, Cain C, Fallahzadeh R, Shahrokni A, Cook D, Ghasemzadeh H. How accurate is your activity tracker? A comparative study of step counts in low-intensity physical activities. JMIR Mhealth Uhealth 2017 Aug 11;5(8):e106 [FREE Full text] [doi: 10.2196/mhealth.6321] [Medline: 28801304]
56. Winslow BD, Nguyen N, Venta KE. Improved mental acuity forecasting with an individualized quantitative sleep model. Front Neurol 2017;8:160 [FREE Full text] [doi: 10.3389/fneur.2017.00160] [Medline: 28487671]
57. Spiotta AM, Fargen KM, Denham SL, Fulton ME, Kellogg R, Young E, et al. Incorporation of a physical education and nutrition program into neurosurgery: a proof of concept pilot program. Neurosurgery 2016 Oct;79(4):613-619. [doi: 10.1227/NEU.0000000000001358] [Medline: 27465847]
58. Modave F, Guo Y, Bian J, Gurka MJ, Parish A, Smith MD, et al. Mobile device accuracy for step counting across age groups. JMIR Mhealth Uhealth 2017 Jun 28;5(6):e88 [FREE Full text] [doi: 10.2196/mhealth.7870] [Medline: 28659255]
59. Dominick GM, Winfree KN, Pohlig RT, Papas MA. Physical activity assessment between consumer- and research-grade accelerometers: a comparative study in free-living conditions. JMIR Mhealth Uhealth 2016 Sep 19;4(3):e110 [FREE Full text] [doi: 10.2196/mhealth.6281] [Medline: 27644334]
60. Schoenfelder E, Moreno M, Wilner M, Whitlock KB, Mendoza JA. Piloting a mobile health intervention to increase physical activity for adolescents with ADHD. Prev Med Rep 2017 Jun;6:210-213 [FREE Full text] [doi: 10.1016/j.pmedr.2017.03.003] [Medline: 28373931]
61. Kooiman TJ, Dontje ML, Sprenger SR, Krijnen WP, van der Schans CP, de Groot M. Reliability and validity of ten consumer activity trackers. BMC Sports Sci Med Rehabil 2015;7:24 [FREE Full text] [doi: 10.1186/s13102-015-0018-5] [Medline: 26464801]
62. Fokkema T, Kooiman TJ, Krijnen WP, van der Schans CP, de Groot M. Reliability and validity of ten consumer activity trackers depend on walking speed. Med Sci Sports Exerc 2017 Apr;49(4):793-800. [doi: 10.1249/MSS.0000000000001146] [Medline: 28319983]
63. Mantua J, Gravel N, Spencer RM. Reliability of sleep measures from four personal health monitoring devices compared to research-based actigraphy and polysomnography. Sensors (Basel) 2016 Dec 05;16(5) [FREE Full text] [doi: 10.3390/s16050646] [Medline: 27164110]
64. Chen JL, Guedes CM, Cooper BA, Lung AE. Short-term efficacy of an innovative mobile phone technology-based intervention for weight management for overweight and obese adolescents: pilot study. Interact J Med Res 2017 Aug 02;6(2):e12 [FREE Full text] [doi: 10.2196/ijmr.7860] [Medline: 28768612]
65. Reichardt LA, Aarden JJ, van Seben R, van der Schaaf M, Engelbert RH, Bosch JA, Hospital-ADL study group. Unravelling the potential mechanisms behind hospitalization-associated disability in older patients; the Hospital-Associated Disability and impact on daily Life (Hospital-ADL) cohort study protocol. BMC Geriatr 2016 Mar 05;16:59 [FREE Full text] [doi: 10.1186/s12877-016-0232-3] [Medline: 26945587]
66. Laranjo L, Lau AY, Martin P, Tong HL, Coiera E. Use of a mobile social networking intervention for weight management: a mixed-methods study protocol. Br Med J Open 2017 Jul 12;7(7):e016665 [FREE Full text] [doi: 10.1136/bmjopen-2017-016665] [Medline: 28706104]
67. Pumper MA, Mendoza JA, Arseniev-Koehler A, Holm M, Waite A, Moreno MA. Using a Facebook group as an adjunct to a pilot mHealth physical activity intervention: a mixed methods approach. Stud Health Technol Inform 2015;219:97-101. [Medline: 26799887]
68. Cook JD, Prairie ML, Plante DT. Utility of the Fitbit Flex to evaluate sleep in major depressive disorder: a comparison against polysomnography and wrist-worn actigraphy. J Affect Disord 2017 Aug 01;217:299-305. [doi: 10.1016/j.jad.2017.04.030] [Medline: 28448949]
69. Jo E, Lewis K, Directo D, Kim MJ, Dolezal BA. Validation of biofeedback wearables for photoplethysmographic heart rate tracking. J Sports Sci Med 2016 Sep;15(3):540-547 [FREE Full text] [Medline: 27803634]
70. Floegel TA, Florez-Pregonero A, Hekler EB, Buman MP. Validation of consumer-based hip and wrist activity monitors in older adults with varied ambulatory abilities. J Gerontol A Biol Sci Med Sci 2016 Jun 02;72(2):229-236. [doi: 10.1093/gerona/glw098] [Medline: 27257217]
71. Alharbi M, Bauman A, Neubeck L, Gallagher R. Validation of Fitbit-Flex as a measure of free-living physical activity in a community-based phase III cardiac rehabilitation population. Eur J Prev Cardiol 2016 Feb 23;23(14):1476-1485. [doi: 10.1177/2047487316634883] [Medline: 26907794]
72. Diaz KM, Krupka DJ, Chang MJ, Shaffer JA, Ma Y, Goldsmith J, et al. Validation of the Fitbit One® for physical activity measurement at an upper torso attachment site. BMC Res Notes 2016 Apr 12;9:213 [FREE Full text] [doi: 10.1186/s13104-016-2020-8] [Medline: 27068022]

73. Sushames A, Edwards A, Thompson F, McDermott R, Gebel K. Validity and reliability of Fitbit Flex for step count, moderate to vigorous physical activity and activity energy expenditure. PLoS One 2016;11(9):e0161224 [FREE Full text] [doi: 10.1371/journal.pone.0161224] [Medline: 27589592]
74. Kang SG, Kang JM, Ko KP, Park SC, Mariani S, Weng J. Validity of a commercial wearable sleep tracker in adult insomnia disorder patients and good sleepers. J Psychosom Res 2017 Jun;97:38-44. [doi: 10.1016/j.jpsychores.2017.03.009] [Medline: 28606497]
75. Voss C, Gardner RF, Dean PH, Harris KC. Validity of commercial activity trackers in children with congenital heart disease. Can J Cardiol 2017 Jun;33(6):799-805. [doi: 10.1016/j.cjca.2016.11.024] [Medline: 28347581]
76. Nelson MB, Kaminsky LA, Dickin DC, Montoye AH. Validity of consumer-based physical activity monitors for specific activity types. Med Sci Sports Exerc 2016 Aug;48(8):1619-1628. [doi: 10.1249/MSS.0000000000000933] [Medline: 27015387]
77. Treacy D, Hassett L, Schurr K, Chagpar S, Paul S, Sherrington C. Validity of different activity monitors to count steps in an inpatient rehabilitation setting. Phys Ther 2017;97(5):581-588.
78. Huang Y, Xu J, Yu B, Shull PB. Validity of FitBit, Jawbone UP, Nike+ and other wearable devices for level and stair walking. Gait Posture 2016;48:36-41. [doi: 10.1016/j.gaitpost.2016.04.025]
79. Gillinov S, Etiwy M, Wang R, Blackburn G, Phelan D, Gillinov AM, et al. Variable accuracy of wearable heart rate monitors during aerobic exercise. Med Sci Sports Exerc 2017 Aug;49(8):1697-1703. [doi: 10.1249/MSS.0000000000001284] [Medline: 28709155]
80. Woodman JA, Crouter SE, Bassett Jr DR, Fitzhugh EC, Boyer WR. Accuracy of consumer monitors for estimating energy expenditure and activity type. Med Sci Sports Exerc 2017;49(2):371-377. [Medline: 27580155]
81. Bronikowski M, Bronikowska M, Glapa A. Do they need goals or support? A report from a goal-setting intervention using physical activity monitors in youth. Int J Environ Res and Public Health 2016;13(9):1-12. [Medline: 27649219]
82. Jones AP, Coombes EG, Griffin SJ, van Sluijs EM. Environmental supportiveness for physical activity in English schoolchildren: a study using Global Positioning Systems. Int J Behav Nutr Phys Act 2009;6:42. [Medline: 19615073]
83. Mooney R, Quinlan LR, Corley G, Godfrey A, Osborough C, Olaighin G. Evaluation of the Finis Swimsense® and the Garmin SwimTM activity monitors for swimming performance and stroke kinematics analysis. PloS One 2017;12(2):e0170902. [Medline: 28178301]
84. An HS, Jones GC, Kang SK, Welk GJ, Lee JM. How valid are wearable physical activity trackers for measuring steps? Eur J Sport Sci 2017 Apr;17(3):360-368. [doi: 10.1080/17461391.2016.1255261] [Medline: 27912681]
85. Ehrler F, Weber C, Lovis C. Influence of pedometer position on pedometer accuracy at various walking speeds: a comparative study. J Med Internet Res 2016 Oct 06;18(10):e268 [FREE Full text] [doi: 10.2196/jmir.5916] [Medline: 27713114]
86. Ammann R, Taube W, Neuhaus M, Wyss T. The influence of the gait-related arm swing on elevation gain measured by sport watches. J Hum Kinet 2016;51:53-60. [doi: 10.1515/hukin-2015-0170]
87. Eyre EL, Duncan MJ, Birch SL, Cox V, Blackett M. Physical activity patterns of ethnic children from low socio-economic environments within the UK. J Sports Sci 2015;33(3):232-242. [Medline: 24998418]
88. Schaffer SD, Holzapfel SD, Fulk G, Bosch PR. Step count accuracy and reliability of two activity tracking devices in people after stroke. Physiother Theory Pract 2017;33(10):788-796. [Medline: 28777710]
89. O'Connell S, ÓLaighin G, Kelly L, Murphy E, Beirne S, Burke N, et al. These shoes are made for walking: sensitivity performance evaluation of commercial activity monitors under the expected conditions and circumstances required to achieve the international daily step goal of 10,000 steps. PLoS One 2016;11(5):e0154956 [FREE Full text] [doi: 10.1371/journal.pone.0154956] [Medline: 27167121]
90. Jennings CA, Berry TR, Carson V, Culos-Reed SN, Duncan MJ, Loitz CC, et al. UWALK: the development of a multi-strategy, community-wide physical activity program. Transl Behav Med 2017;7(1):16-27. [doi: 10.1007/s13142-016-0417-5]
91. Price K, Bird SR, Lythgo N, Raj IS, Wong JY, Lynch C. Validation of the Fitbit One, Garmin Vivofit and Jawbone UP activity tracker in estimation of energy expenditure during treadmill walking and running. J Med Eng Technol 2017;41(3):208-215. [Medline: 27919170]
92. Claes J, Buys R, Avila A, Finlay D, Kennedy A, Guldenring D, et al. Validity of heart rate measurements by the Garmin Forerunner 225 at different walking intensities. J Med Eng Technol 2017;41(6):480-485. [doi: 10.1080/03091902.2017.1333166]
93. O'Brien A, Schlosser RW, Shane HC, Abramson J, Allen AA, Flynn S, et al. Brief report: just-in-time visual supports to children with autism via the Apple watch®: a pilot feasibility study. J Autism Dev Disord 2016 Dec;46(12):3818-3823. [doi: 10.1007/s10803-016-2891-5] [Medline: 27573856]
94. Cook S, Stauffer JC, Goy JJ, Graf D, Puricel S, Frobert A, et al. Heart rate never lies: interventional cardiologist and Braude's quote revised. Open Heart 2016;3(1):e000373 [FREE Full text] [doi: 10.1136/openhrt-2015-000373] [Medline: 26835145]
95. Mercer K, Giangregorio L, Schneider E, Chilana P, Li M, Grindrod K. Acceptance of commercially available wearable activity trackers among adults aged over 50 and with chronic illness: a mixed-methods evaluation. JMIR mHealth uHealth 2016;4(1):e7 [FREE Full text] [doi: 10.2196/mhealth.4225] [Medline: 26818775]

96. El-Amrawy F, Nounou MI. Are currently available wearable devices for activity tracking and heart rate monitoring accurate, precise, and medically beneficial? Healthc Inform Res 2015 Oct;21(4):315-320 [FREE Full text] [doi: 10.4258/hir.2015.21.4.315] [Medline: 26618039]
97. Ferguson T, Rowlands AV, Olds T, Maher C. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. Int J Behav Nutr Phys Act 2015;12:42 [FREE Full text] [doi: 10.1186/s12966-015-0201-9] [Medline: 25890168]
98. Hernández-Vicente A, Santos-Lozano A, De Cocker K, Garatachea N. Validation study of Polar V800 accelerometer. Ann Transl Med 2016 Aug;4(15):278 [FREE Full text] [doi: 10.21037/atm.2016.07.16] [Medline: 27570772]
99. Giles D, Draper N, Neil W. Validity of the Polar V800 heart rate monitor to measure RR intervals at rest. Eur J Appl Physiol 2016 Mar;116(3):563-571 [FREE Full text] [doi: 10.1007/s00421-015-3303-9] [Medline: 26708360]
100. Gruwez A, Libert W, Ameye L, Bruyneel M. Reliability of commercially available sleep and activity trackers with manual switch-to-sleep mode activation in free-living healthy individuals. Int J Med Inform 2017 Jun;102:87-92. [doi: 10.1016/j.ijmedinf.2017.03.008] [Medline: 28495352]
101. O'Connell S, ÓLaighin G, Quinlan LR. When a step is not a step! Specificity analysis of five physical activity monitors. PLoS One 2017;12(1):e0169616 [FREE Full text] [doi: 10.1371/journal.pone.0169616] [Medline: 28085918]
102. Parak J, Korhonen I. Evaluation of wearable consumer heart rate monitors based on photopletysmography. Conf Proc IEEE Eng Med Biol Soc 2014;2014:3670-3673. [doi: 10.1109/EMBC.2014.6944419] [Medline: 25570787]
103. Spierer DK, Rosen Z, Litman LL, Fujii K. Validation of photoplethysmography as a method to detect heart rate during rest and exercise. J Med Eng Technol 2015;39(5):264-271. [doi: 10.3109/03091902.2015.1047536] [Medline: 26112379]
104. Parak J, Uuskoski M, Machek J, Korhonen I. Estimating heart rate, energy expenditure, and physical performance with a wrist photoplethysmographic device during running. JMIR mHealth uHealth 2017 Jul 25;5(7):e97 [FREE Full text] [doi: 10.2196/mhealth.7437] [Medline: 28743682]
105. Delgado-Gonzalo R, Parak J, Tarniceriu A, Renevey P, Bertschi M, Korhonen I. Evaluation of accuracy and reliability of PulseOn optical heart rate monitoring device. Conf Proc IEEE Eng Med Biol Soc 2015 Aug;2015:430-433. [doi: 10.1109/EMBC.2015.7318391] [Medline: 26736291]
106. Parak J, Tarniceriu A, Renevey P, Bertschi M, Delgado-Gonzalo R, Korhonen I. Evaluation of the beat-to-beat detection accuracy of PulseOn wearable optical heart rate monitor. Conf Proc IEEE Eng Med Biol Soc 2015 Aug;2015:8099-8102. [doi: 10.1109/EMBC.2015.7320273] [Medline: 26738173]
107. Apple. 2017. WatchOS - Apple Developer URL: https://developer.apple.com/watchos [accessed 2017-09-19] [WebCite Cache ID 6tamfJinU]
108. Apple. 2017. HealthKit - Apple Developer URL: https://developer.apple.com/healthkit/ [accessed 2017-09-13] [WebCite Cache ID 6tRfiSKVw]
109. Fitbit. 2017. Fitbit Developer URL: https://dev.fitbit.com/ [accessed 2017-09-22] [WebCite Cache ID 6tfGRUrmR]
110. Garmin. Garmin Developers URL: https://developer.garmin.com/ [accessed 2017-09-13] [WebCite Cache ID 6tRfk6phg]
111. Misfit. Misfit Developer Toolkit URL: https://build.misfit.com/ [accessed 2017-09-13] [WebCite Cache ID 6tRgRKOAN]
112. Polar. Polar AccessLink URL: https://www.polar.com/en/connect_with_polar/polar_accesslink [accessed 2017-09-19] [WebCite Cache ID 6tamdLu6Q]
113. Samsung. Samsung Health - Samsung Developers URL: http://developer.samsung.com/health [accessed 2017-09-13] [WebCite Cache ID 6tRftLAHj]
114. TomTom. TomTom Developer Portal URL: https://developer.tomtom.com/ [accessed 2017-09-13] [WebCite Cache ID 6tRgHGW0A]
115. Nokia. Nokia Health API Developer Documentation URL: https://developer.health.nokia.com/api [accessed 2017-09-13] [WebCite Cache ID 6tRgHmHhR]
116. PR Newswire Association. Fossil Group Reimagines the Watch URL: https://www.prnewswire.com/news-releases/fossil-group-reimagines-the-watch-300427943.html [accessed 2018-03-01] [WebCite Cache ID 6xa1Z8fI6]

Abbreviations

API: application programming interface
EE: energy expenditure
GPS: global positioning system
HR: heart rate
IMU: inertial measurement unit
MEDLINE: Medical Literature Analysis and Retrieval System Online
PA: physical activity
PPG: photoplethysmography
SDK: software development kit

Edited by G Eysenbach; submitted 17.11.17; peer-reviewed by J Sanders, P Wark, K Winfree, R Fallahzadeh, C Fernández; comments to author 07.12.17; revised version received 18.12.17; accepted 06.01.18; published 22.03.18.

Please cite as:
Henriksen A, Haugen Mikalsen M, Woldaregay AZ, Muzny M, Hartvigsen G, Hopstock LA, Grimsgaard S
Using Fitness Trackers and Smartwatches to Measure Physical Activity in Research: Analysis of Consumer Wrist-Worn Wearables
J Med Internet Res 2018;20(3):e110
URL: http://www.jmir.org/2018/3/e110/
doi:10.2196/jmir.9157
PMID:29567635

Original Paper

Integration of a Technology-Based Mental Health Screening Program Into Routine Practices of Primary Health Care Services in Peru (The Allillanchu Project): Development and Implementation

Francisco Diez-Canseco[1], MPH; Mauricio Toyama[1], BA; Alessandra Ipince[1], MA; Silvana Perez-Leon[1], BA; Victoria Cavero[1], BA; Ricardo Araya[2], MD, PhD; J Jaime Miranda[1,3], MD, PhD

[1]CRONICAS Center of Excellence in Chronic Diseases, Universidad Peruana Cayetano Heredia, Lima, Peru

[2]Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom

[3]School of Medicine, Universidad Peruana Cayetano Heredia, Lima, Peru

Corresponding Author:
Francisco Diez-Canseco, MPH
CRONICAS Center of Excellence in Chronic Diseases
Universidad Peruana Cayetano Heredia
Av. Armendariz 497
Miraflores
Lima
Peru
Phone: 51 12416978
Email: fdiezcanseco@gmail.com

Abstract

Background: Despite their high prevalence and significant burden, mental disorders such as depression remain largely underdiagnosed and undertreated.

Objective: The aim of the *Allillanchu Project* was to design, develop, and test an intervention to promote early detection, opportune referral, and access to treatment of patients with mental disorders attending public primary health care (PHC) services in Lima, Peru.

Methods: The project had a multiphase design: formative study, development of intervention components, and implementation. The intervention combined three strategies: training of PHC providers (PHCPs), task shifting the detection and referral of mental disorders, and a mobile health (mHealth) component comprising a screening app followed by motivational and reminder short message service (SMS) to identify at-risk patients. The intervention was implemented by 22 PHCPs from five health centers, working in antenatal care, tuberculosis, chronic diseases, and HIV or AIDS services.

Results: Over a period of 9 weeks, from September 2015 to November 2015, 733 patients were screened by the 22 PHCPs during routine consultations, and 762 screening were completed in total. The chronic diseases (49.9%, 380/762) and antenatal care services (36.7%, 380/762) had the higher number of screenings. Time constraints and workload were the main barriers to implementing the screening, whereas the use of technology, training, and supervision of the PHCPs by the research team were identified as facilitators. Of the 733 patients, 21.7% (159/733) screened positively and were advised to seek specialized care. Out of the 159 patients with a positive screening result, 127 had a follow-up interview, 72.4% (92/127) reported seeking specialized care, and 55.1% (70/127) stated seeing a specialist. Both patients and PHCPs recognized the utility of the screening and identified some key challenges to its wider implementation.

Conclusions: The use of a screening app supported by training and supervision is feasible and uncovers a high prevalence of unidentified psychological symptoms in primary care. To increase its sustainability and utility, this procedure can be incorporated into the routine practices of existing health care services, following tailoring to the resources and features of each service. The early detection of psychological symptoms by a PHCP within a regular consultation, followed by adequate advice and support, can lead to a significant percentage of patients accessing specialized care and reducing the treatment gap of mental disorders.

(J Med Internet Res 2018;20(3):e100) doi:10.2196/jmir.9208

KEYWORDS

mental health; mHealth; SMS; textmessaging; screening; mobile health; health services research

Introduction

The Comorbidity of Mental Disorders and Physical Conditions

Worldwide, mental health disorders are highly prevalent and disabling conditions [1]. In Peru, mental disorders affect 1 in 5 people and are the leading cause of disease burden [2,3]. Certain population subgroups experience a higher burden of mental disorders, such as primary health care (PHC) services users. For instance, in Peru, the prevalence of depression is as high as 50% among patients with tuberculosis [2], 40% during pregnancy [3-5], 68% among female patients living with HIV or AIDS [6], and up to 57.8% for patients with diabetes [7-9]. In addition to this high comorbidity, for people living with a physical disease, depression has been associated with reduced treatment adherence, poorer prognosis, greater disability, and higher mortality [10-12]. Among pregnant women, depression is associated with underutilization of antenatal care services, premature birth, lower birth weight, and constitutes the main risk factor for postpartum depression [13-16]. Despite the high prevalence and negative impact on patients' lives, depression and other mental disorders go largely underdiagnosed and undertreated [2,6]. Indeed, up to 85% of people reporting a need for mental health care declared not receiving any care [17,18].

Strategies to Address the Comorbidity

Addressing comorbid physical and mental conditions, particularly within primary care, provides an opportunity to reduce the existing mental health treatment gap [19]. Accordingly, the World Health Organization proposes the integration of mental health services with general health care as a key solution to this gap, even in low- and middle-income countries (LMICs) [20]. Furthermore, there is growing evidence of the effectiveness of task shifting in the management of some chronic conditions, including mental disorders [21].

Mobile communications have also shown increasing potential for the improvement of health care [22]. Some mobile health (mHealth) strategies have proven to be useful in improving the efficiency of health care delivery, including interventions provided via technological platforms for the management of physical conditions [23-25], mental disorders [26-28], and as clinical decision support tools [29-32]. Moreover, strategies using short message service (SMS) as periodic reminders and informative messages have had a positive impact on medication adherence, appointment attendance, symptom monitoring, satisfaction with health services, and promotion of healthy behavior [27,33,34].

Packaging and deploying pragmatic strategies to introduce mental health screening and treatment within existing PHC platforms are needed in Peru and similar LMIC settings [35]. In this paper, we describe the design, development, and implementation of the *Allillanchu Project* and discuss the feasibility and challenges to its implementation at a larger scale.

Methods

Study Description

The *Allillanchu Project* aimed to design, develop, and test an intervention to promote early detection, opportune referral, and access to treatment of patients with common mental disorders attending public PHC services in Lima, Peru. *Allillanchu*, meaning *How are you? How are you feeling?* in *Quechua*, a Peruvian indigenous language, was chosen to emphasize the need to integrate mental health care into PHC providers' (PHCPs') routine practices. The intervention combines three key strategies: (1) training of PHCPs in the use of a screening tool to detect and manage depression, anxiety, psychosis, convulsive disorder, and alcoholism; (2) task shifting of detection and referral of mental disorders to PHCPs; and (3) an mHealth component that comprises a screening app plus motivational and reminder SMS text messages to patients.

Study Design

The project comprised three phases: phase 1, preintervention activities; phase 2, design of intervention components; and phase 3, implementation (see *Procedures*). In each of these phases, we collected quantitative and qualitative information. This project followed a mixed-methods approach, through a multiphase design [36]. This methodology is commonly used with feasibility studies. This design included both sequential and concurrent data collection. During phases 1 and 2, data were collected sequentially; and in phase 3, qualitative and quantitative data were concurrently collected (see Figure 1).

Setting

In Peru, there are two main public health systems: the Ministry of Health and the social security system (EsSalud). The intervention was deployed in five different public PHC centers: three from the Ministry of Health and two from EsSalud. All of them served low-income populations living in the northern districts of Lima, Peru's capital. The study involved PHCPs and patients from these five health centers that included at least one of the following health services (see Table 1): antenatal control (3 services), tuberculosis (4 services), chronic diseases (2 services), and HIV or AIDS (1 service). Each health service had between one and four PHCPs, each attending 15 to 20 patients per day, in 6-hour working shifts.

All participating health centers had at least one psychologist as staff; the only specialized mental health professionals available at the primary care level. Psychiatry specialists work in general and psychiatric hospitals; and patients had to be referred from PHC to be able to access these professionals.

Furthermore, the health centers from the social security system have a service of complementary medicine (ie, Tai chi) for referred patients to improve their physical health condition or their emotional well-being.

Figure 1. Collection of data in the study phases. SMS: short message service; PHCP: primary health care provider.

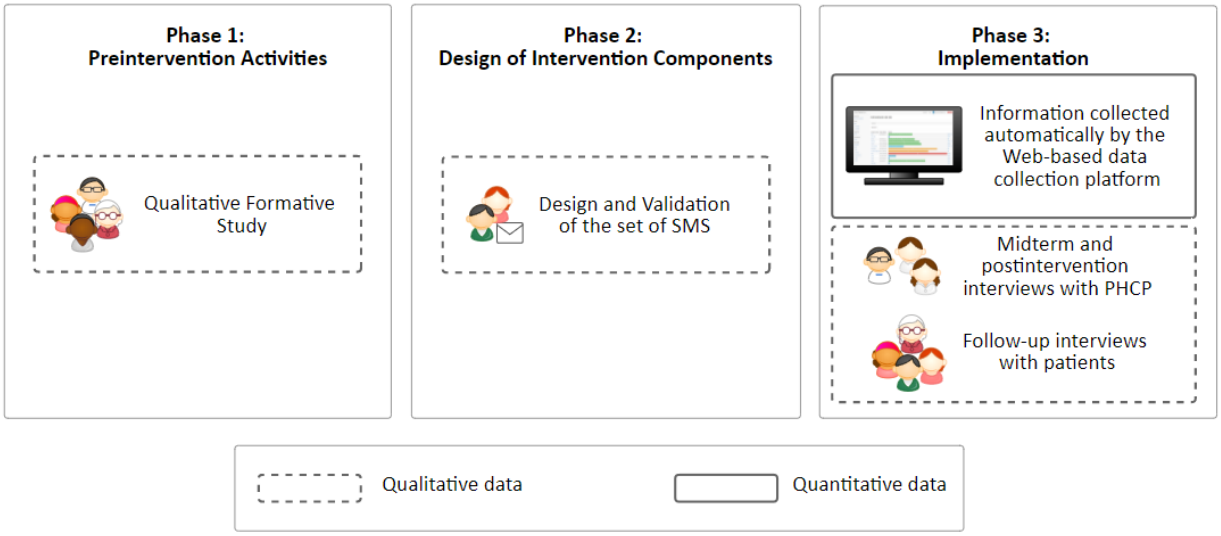


Table 1. Number of health centers and services included in the study.

Services	Ministry of Health			Social Security System (EsSalud)		Total
	Health center 1	Health center 2	Health center 3	Health center 4	Health center 5	
Tuberculosis	0	1	1	1	1	4
Antenatal care	1	1	0	1	0	3
Chronic diseases	0	0	0	1	1	2
HIV or AIDS	0	0	1	0	0	1
Total	1	2	2	3	2	10

Participants

Primary Health Care Providers

A total of 22 PHCPs (12 midwives, 8 nurses, and 2 nurse assistants) out of 29 working in these facilities (76%) agreed to include the screening and the referral of positive cases as part of their routines during the 9-week study period.

Patients

Adult patients, aged ≥18 years and attending participating services, were invited into the study and had to consent to participate in the study. Participants who screened positive were recommended by the PHCPs to seek mental health care, received a set of tailored SMS text messages to motivate them to seek such care, and were contacted by the research team for a follow-up assessment (see *Phase 3: Implementation*).

Procedures

The *Allillanchu Project* involved three different phases (see Figure 2):

1. Preintervention activities: explored the implementation setting to develop a suitable and context-specific intervention
2. Design of intervention components: developed the key components needed to deliver the intervention, including the training and supervision of PHCPs
3. Implementation: tested the integration of the mental health screening and referral in primary care combining the mHealth strategies, task shifting, and training of PHCPs.

Phase 1: Preintervention Activities

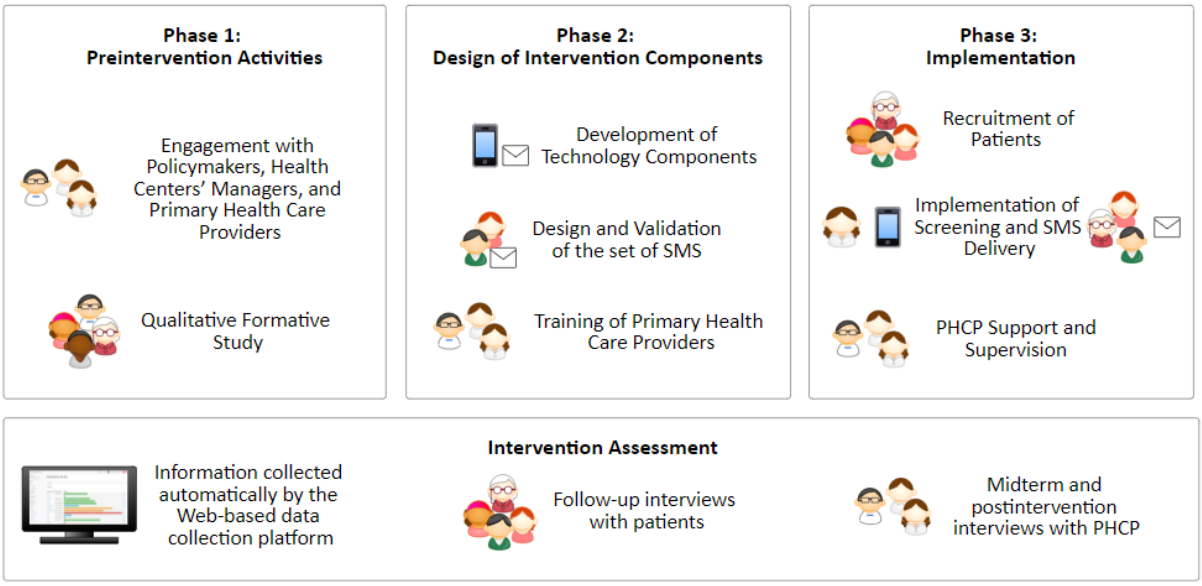
Engagement of Policy Makers, Health Centers' Managers, and Primary Health Care Providers

This stage consisted of a year-long period of meetings with policy makers and the personnel at the health centers, including the centers' directors (n=5), heads of the services (n=10), and PHCPs (n=29). These meetings aimed to present the study, assess its feasibility, and secure the buy-in of managers to obtain the necessary approvals to implement the study in their facilities.

Qualitative Formative Study

From September 2014 to December 2014, we conducted a qualitative study to identify implementation barriers and facilitators. The research team interviewed patients and personnel of health centers where the intervention will later be implemented, including 22 PHCPs, four clinical psychologists and 37 patients. The interviews were conducted using semistructured guides, and the duration was approximately of 45 min. All interviews were recorded, transcribed verbatim, and analyzed by the research team using Atlas.Ti, version 7 (Scientific Software Development GmbH).

Figure 2. Allillanchu project's phases. SMS: short message service; PHCP: primary health care provider.



The main results of this qualitative formative study were as follows: (1) screening for mental disorders was not a quotidian practice in the services, (2) PHCPs were willing to introduce the screening in their routines but anticipated time constraints, (3) though it was unusual for them to do so, many patients were willing to seek mental health care if they were advised to do so, (4) mental health specialists were scarce in participating health centers, and (5) interviewed clinical psychologists recognized the value of introducing a mental disorders' screening in primary care but were concerned about PHCP's ability and willingness to do so. Findings of the qualitative study will be reported in detail elsewhere.

Phase 2: Design of Intervention Components

We proposed the inclusion of a technology-based screening tool into the routines of trained PHCPs, complemented by a set of automated SMS text messages to encourage patients who screen positive to seek specialized care and to advise on where to seek help.

Development of Technology Components

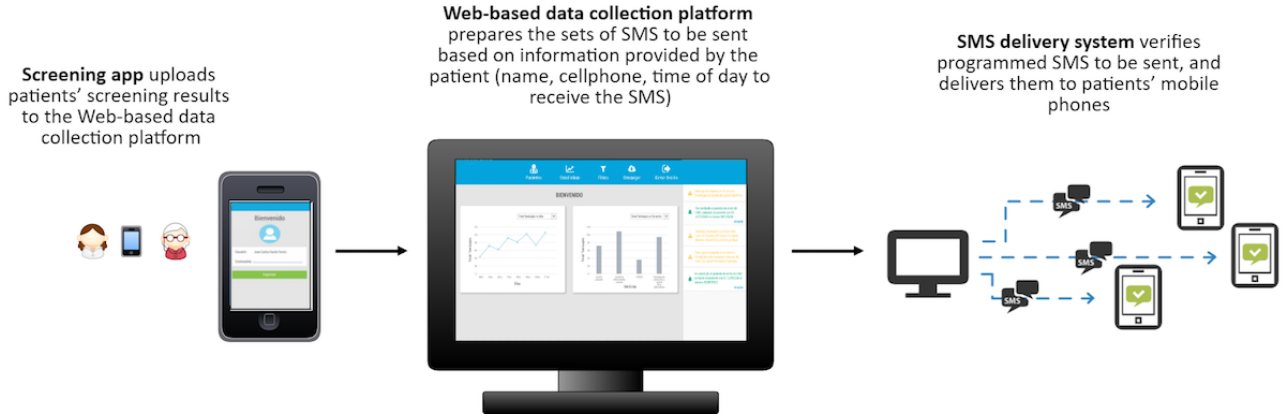
The technology consisted of three integrated components: a screening app, a Web-based data collection platform, and an automated SMS text message delivery app.

The *screening app*, installed on a tablet, used the Self Report Questionnaire (SRQ). The guiding principle was to develop an app to (1) Produce immediate results on symptomatology that may require professional help, (2) Offer guidelines to PHCPs to advise the patient on how and where to seek help, and (3) Ask PHCPs to report on actions taken with positive cases (ie, recommendation to attend to a psychology service). The SRQ [37], a screening tool developed by the World Health Organization, is recommended in the current mental health reform in Peru as the tool to be used in PHC. The SRQ version used by the Peruvian Ministry of Health consists of 28 yes or no questions, 18 of which screen for depression and anxiety, four for psychosis, one for convulsive disorder, and five questions for alcoholism (see Multimedia Appendix 1). The Peruvian social security system was planning to implement a shorter version of the SRQ in their services at the time; therefore, these centers used a short version, which includes the first 18 items assessing depression and anxiety. The team further added two questions to assess suicide risk. These questions were asked only when patients answered positively to question 18 of the SRQ, which establishes the presence of suicidal ideation. Each PHCP had a personal log-in that displayed the SRQ version for their health system, Ministry of Health, or EsSalud.

We collected user feedback on the screening app through a small pilot with five PHCPs from different health services (n=4) and centers (n=3) for 2 weeks. The pilot found that PHCPs had no difficulties using the screening app; they found it easy to use and were able to include it in their daily routines. In the 2-week pilot, PHCPs screened a total of 47 patients.

The *Web-based data collection platform* stored all the information collected by PHCPs using the screening app, allowing the research team to monitor progress in real time (see *Intervention Assessment*). As patients were screened, the app uploaded the results to the Web-based data collection platform in real time using a mobile data connection. Furthermore, the information provided by the patient during the recruitment process was also uploaded to the Web platform (name, cellphone number, and time of day to receive the SMS text message) and consolidated under a patient ID. Thus, based on the screening results, the Web platform was able to match a positive screening under the specified ID with the patient's personal data to automatically send an SMS text message (see Figure 3).

Figure 3. Integration of technology components. SMS: short message service.



Design and Validation of the Set of Short Message Service

The research team developed a set of five SMS text messages to further motivate patients to seek mental health care. Patients who screened positive and were referred to specialized care received three SMS text messages per week, during 2 weeks. Of the five SMS text messages, sent every other day, one aimed to remind patients where and when to find mental health care (*reminder SMS text message*), whereas the other four sought to motivate patients to seek help (*motivational SMS text message*) addressing either barriers or facilitators for accessing mental health care. The pattern of SMS text message delivery over the 2 weeks was the same: day 1: reminder SMS text message, day 2: motivational SMS text message, and day 3: motivational SMS text message. As the reminder SMS text message was sent twice, each patient received a total of six SMS text messages over the 2-week period.

All SMS text messages were designed and validated through two pilot studies, conducted from December 2014 to January 2015, with 63 patients with similar characteristics to those involved in the implementation phase. The first pilot study aimed to explore (1) Patients' use of mobile phones and SMS text messages and (2) Barriers and facilitators for seeking mental health care. On the basis of these results, we developed seven SMS text messages that were validated and ranked by patients in a second study [38]. This resulted in the selection of the set of five SMS text messages for the implementation phase that (1) Were simple and direct, (2) Avoided mentioning health conditions (eg, depression), (3) Were tailored with the patient's name and health centers' name as signature, and (4) Addressed the barriers (ie, lack of time and money) and highlighted the positive effects of looking for specialized mental health care (ie, feeling better, having someone to talk to, and receive guidance), as reported by the interviewed patients.

Training of Primary Health Care Providers

To participate in the implementation, the 22 PHCPs were trained in the use of the screening app. This training consisted of two sessions of 8.5 hours each, which were developed and offered by the project team, which included psychologists, one sociologist, and a consultant with a background in psychotherapy. This activity sought to offer knowledge, skills, and motivation to PHCPs to enable the routine screening and referral of patients to available mental health services as indicated by the app. The mental health content was based on the mhGAP training modules [39]. It included presentations, role-playing, and practical exercises.

Phase 3: Implementation

Recruitment of Patients

From September 2015 to November of 2015, a team of 10 recruiters enrolled patients (see *Description of Study Participants*). The recruitment took place in the waiting rooms before the patients' appointments and involved them providing informed consent to be part of the study and providing personal data to be uploaded to the Web-based data collection system. The informed consent explained that in the case of a positive screening during one of their regular consultations, they would receive SMS text messages and have a follow-up interview with the research team.

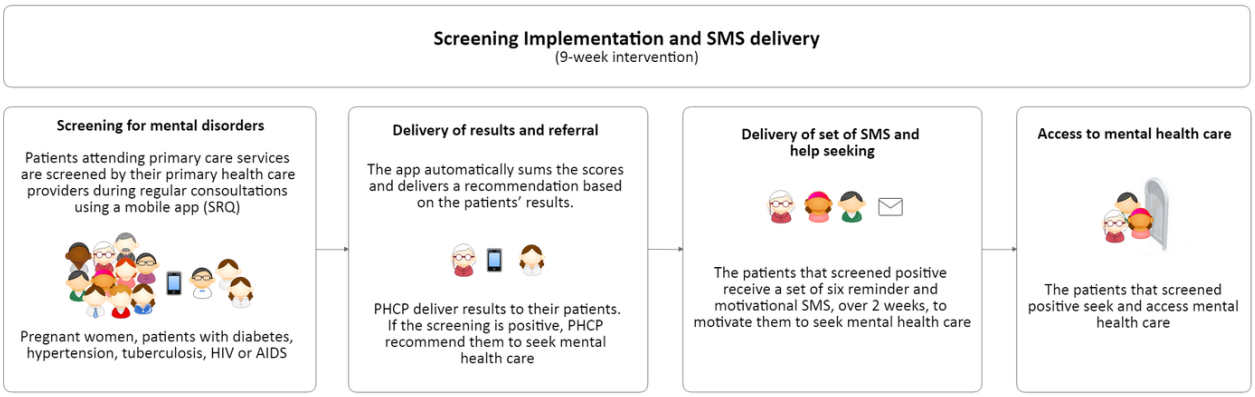
Screening Implementation and Short Message Service Delivery

The 22 trained PHCPs were invited to use the screening app in their regular consultations during 9 weeks, from 2015 September to November of 2015, and to refer patients with a positive result to mental health care—either psychology service, complementary medicine, or general medicine—according to each center's protocol. The tablets were provided by the *Allillanchu Project* and included a mobile data plan to upload the information to the Web-based data collection platform in real time (see Figure 4).

Primary Health Care Provider's Supervision and Support

Over the 9 weeks, the research team provided telephone and face-to-face support and supervision to the PHCPs to ensure the correct implementation of the screening. The Web-based data collection platform was monitored on a daily basis to identify PHCPs who were not using the app as intended and to contact them to assess and solve the situation.

Figure 4. Intervention procedures. SRQ: Self-Report Questionnaire; PHCP: primary health care provider.



All of the PHCPs had supervision meetings with the research team at least twice a week. In these meetings, support was offered by solving any issues or doubts regarding the use of the tablet or the app and by accompanying the PHCPs in managing potentially difficult situations. Problems with the administration of the SRQ were also assessed and solved. PHCPs could initiate contact with the research team through a "help" function installed in the app or a phone call.

Intervention Assessment

To assess if and how the intervention promoted early detection, opportune referral, and access to treatment of patients with mental disorders, we used the following methods and sources of information:

1. *Information collected automatically by the Web-based data collection platform.* The platform collected and provided information on the screening progress by health center, service, PHCP, patient, and over time, including the number of screenings (completed and incomplete), number of positive cases by type of disorder and of suicide risk, and actions taken by the health provider with each patient (eg, referral to the psychologist). It also included demographic data of all the screened patients.
2. *Follow-up interviews with patients* (Multimedia Appendix 2) . We conducted face-to-face structured postintervention interviews with patients. The interview explored if the patient had sought and received mental health care after being referred by the PHCP, their motivation to seek—or not—specialized care and how the SMS text message encouraged them—or not—to do so, as well as perceived barriers and their opinion on received care. The research team interviewed patients 3 weeks after they screened positive.
3. *Midterm and postintervention interviews with PHCPs* (Multimedia Appendix 3) . Face-to-face semistructured interviews were conducted halfway through and at the end of the intervention. The midterm evaluation assessed PHCPs' personal experiences with the screening (eg, use of technology and delivering results to patients) and aimed to identify and solve problems. The postintervention evaluation explored the experiences and opinions of participating PHCPs, to identify barriers and receive suggestions for improvement, and to assess the perceived feasibility and willingness to continue implementing the screening.

Data Analysis

The screening and referral of patients' data, stored in the Web-based data collection platform, was exported and analyzed using Statistical Package for the Social Sciences (SPSS) statistics for Windows, version 23.0 (IBM Corp) through descriptive analyses, by reporting frequencies and percentages.

Information from patients' and PHCP's interviews was registered and analyzed in a similar way. All interviews were audiorecorded, transcribed, and summarized in a matrix. The qualitative information was organized by participant (rows) and by themes of each interview guide (columns) [40] (Multimedia Appendix 4). Each theme included a summary-cell of interviewee experiences or opinions, as well as quotations. Data were organized and coded by three members of the research team after an initial period of standardization of criteria and practice. The answers to closed-ended questions (eg, "Did you seek care from a general practitioner or psychologist as recommended?") were analyzed, reporting frequencies and percentages.

Ethics

The study protocol, informed consent forms, and instruments were approved by the institutional review board of the Universidad Peruana Cayetano Heredia. The study was also approved by the directors of the five participating PHC centers. Participant PHCPs signed a consent form that described the project goals and procedures, as well as their commitment, the benefits and incentives of participation (which comprised a training-assistance certificate), and funding support toward a one credit brief course of their choice, for up to US $30, at the end of the study.

Patients provided oral informed consent on the following components, which would be implemented only in the case of positive screening results during one of their regular consultations: (1) receive six SMS text messages over 2 weeks, (2) have a follow-up interview, and (3) consent for the interview to be recorded. Ethical approval was also obtained for collection of personal data (name and cellphone number) during the design and validation of the set of SMS text messages.

Results

Description of Study Participants

A total of 22 PHCPs (12 midwives, 8 nurses, and 2 nurse assistants), all female, participated in the study. During recruitment, 2580 patients (22.4% men) attending the selected facilities were invited to participate in the study, and 1772 (68.7%) accepted (see Table 2). The average age of enrolled participants was 42.8 years (SD 20.3).

Implementation of the Screening

The PHCPs performed 762 screenings with 733 patients, as 29 of these were screened for a second time in a following consultation. During the intervention period, the average number of screenings per week was 85 (SD 38). However, the number of screenings per week varied and tended to progressively decrease, having its highest and lowest points during the second (159 screenings) and last week of the intervention (52 screenings), respectively.

Chronic disease services had the highest number of screenings, accounting for almost half of the 762 screenings (49.9%, 380/762), followed by the antenatal care services (36.7%, 280/762), the tuberculosis services (12.1%, 92/762) and the HIV or AIDS services (1.3%, 10/762), respectively.

Positive Cases Detection and Referrals

Out of the 733 patients screened, 159 of them (21.7%) had a positive result by the SRQ. Of those 159, 150 (94.3%) screened positive for one disorder and 9 (5.7%) for two disorders.

Distributed by health service, the highest prevalence was found in the HIV or AIDS service according to the SRQ, where 7 out of 10 screened patients were positive (70%), followed by the chronic diseases services (90/364, 24.7%), antenatal care services (49/274, 17.9%), and tuberculosis services (13/85, 15.3%).

The most prevalent mental disorders were depression or anxiety, grouped as a single condition in the SRQ, with 125/733 patients screening positive for at least one of these conditions (17.1%). At the Ministry of Health centers, who implemented the 28-item version of the SRQ, 41/194 screened patients (21.1%) answered positively to at least one of the four items for psychosis, and 2 (1%) were positive for alcoholism. No cases of convulsion were detected. Additionally, 22 of all 733 screened patients (3%) reported a current suicidal ideation.

According to the PHCPs' reports, the vast majority of the 159 cases detected were referred to mental health care. These referrals consisted in advising their patients to seek specialized care at their health center. Of these 159 patients, 107 (67.3%) were referred to the psychology service; 27 (17%) of the cases were accompanied by the PHCP to a health services (psychology, general medicine) to facilitate access; 4 (2.5%) of the suggestions involved seeking care at a different service (complementary medicine, general medicine); 11 (6.9%) were patients already receiving specialized care and were advised to continue their treatment; 6 (3.8%) reported other actions, that is, providing a piece of paper with the health service of referral written and the signature of the PHCP; and 4 (2.5%) omitted reporting the actions taken.

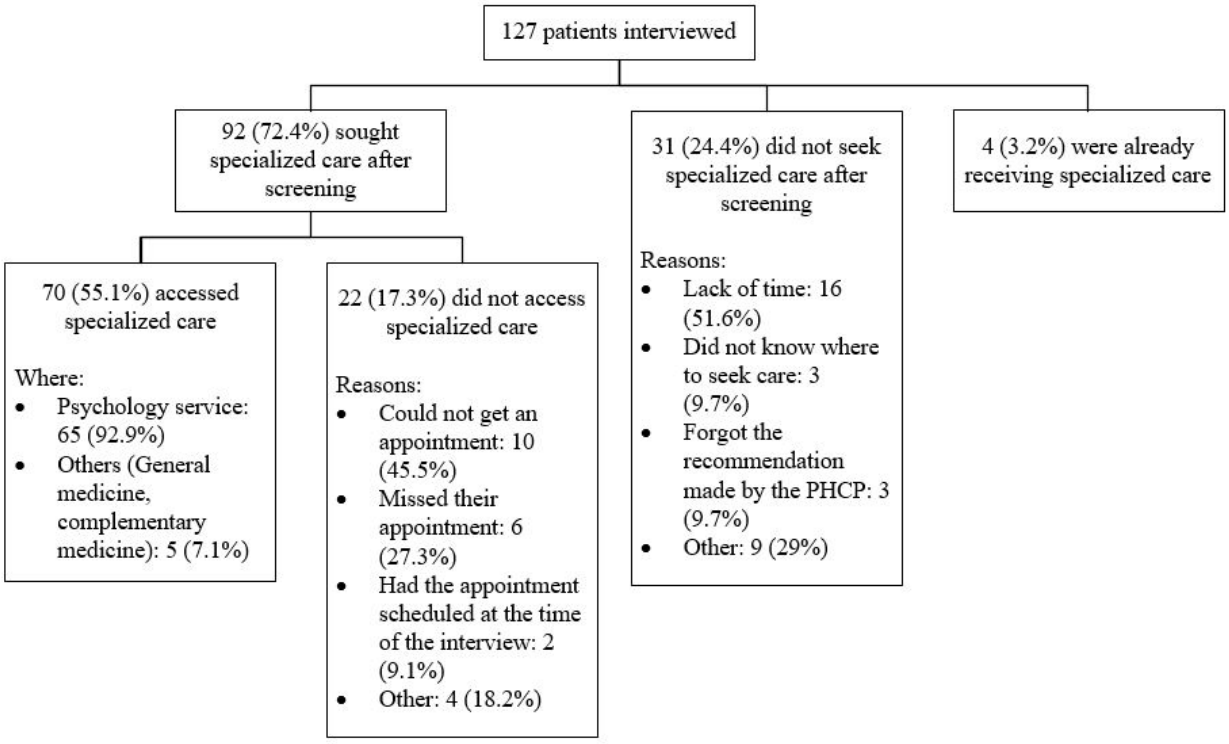
Help-Seeking and Access to Mental Health Care

Out of the 159 patients with a positive screening result, 143 previously gave their consent to participate in the follow-up interview, and 127/143 patients (88.8%) were interviewed. The average age of interviewed participants was 49 years (SD 19), 101/127 (79.5%) were women, and 92/127 (72.4%) reported having sought specialized care after receiving the advice of the PHCP. In addition, 70/127 (55.1%) accessed care, having had at least one consultation in a specialized service after the screening (see Figure 5).

Table 2. Characteristics of enrolled patients.

Characteristics	n (%)
Sex	
Male	346 (19.5)
Female	1426 (80.5)
Health system	
Ministry of Health	609 (34.4)
Social security system (EsSalud)	1163 (65.6)
Health service	
Antenatal care service	931 (52.5)
Chronic diseases service	709 (40)
Tuberculosis service	121 (6.8)
HIV or AIDS service	11 (0.6)

Figure 5. Breakdown of patients interviewed.



All patients who screened positive received a total of six SMS text messages over 2 weeks. However, only 35/127 interviewees (28%) were eligible to answer questions about the perceived effects of the SMS text message, as many participants had already sought specialized care before receiving the first SMS text message, therefore defeating their motivational purpose. Others did not seek specialized care, did not consent to receive the SMS text message, or reported having problems with their cellphone. Of the 35 patients who were suitable to discuss the perceived effects of receiving the SMS text messages, 30/35 (85.7%) considered the set of SMS text messages to be effective in promoting motivation to seek care.

During the interviews, the patients tended to focus on the difficulties and symptoms they were facing, but also offered points of view that further support the implementation of the screening and showcase the benefits of accessing a mental health specialist, as illustrated in the following quotes:

> *I think it is very good that midwives ask their patients how they feel [...] for example, when I came I was going through a deep depression and I wasn't accepting my pregnancy. After the midwife asked me some questions, she sent me to the psychologist of the health center and he instantly saw me. We had a conversation of at least an hour and a half, he helped me a lot. Thanks to them, my depression has decreased, now I feel better, calmer.* [Patient of antenatal care, female, 29 years]

> *The psychologist treated me really well, he made me notice some things and told me it was necessary that my wife come to see him as well, because most of the problems I had were due to our conflictive relationship. Now I feel much better, I am trying to do my best, I am trying to stop having negative thoughts, of eliminating me [...] I want to have another appointment with the psychologist, because with the help of a professional I can receive guidance to think in a different way.* [Patient of the chronic diseases service, male, 64 years]

Primary Health Care Provider's Assessment of the Project

Of the 22 PHCPs, 21 were interviewed for the midterm evaluation and 22 for the final evaluation.

Training

As part of the training, a pre- and posttest evaluation was conducted with a maximum possible score of 20 on each test. The average score at pretest was 12.1 (SD 2.3) and 16.5 (SD 1.35) at posttest, showing an increase of 4.4 points toward the end of training.

In addition, the participating PHCPs assessed the training at the end of the sessions and after the 9-week intervention. They believed the training made them more aware of the importance of caring for people's mental health, provided them with knowledge and skills in mental health—an understudied topic in their professional education—taught them how to use a screening tool to assess their patients' mental health, and offered them necessary skills to help their patients. However, almost half of the PHCPs suggested more time to be allocated to discussing the questions of the SRQ, the screening tool.

Screening

When asking to assess their ability to implement the task as intended, the answers showed a great variability among the PHCPs of the different health services involved. Specifically,

most of the midwives from the antenatal care services reported not being able to implement the screening as part of their routine because of the high workload and the great amount of paperwork for their current tasks, as illustrated in the following quote:

> *Some days I was screening and I was really excited to do it, but when I was with the third pregnant woman screened, I looked at my watch and it was already 10:30 am, and by noon I had to see 12 women. That meant that I had to stop using the tablet and rush to finish on time with all 12 consultations from my shift.* [Midwife, antenatal care service]

Nurses from the chronic disease service reported being unable to do the screening when conducting the weekly educational workshop on disease management with their patients. Conversely, the tuberculosis service's nurses were able to schedule the screening with their patients. One nurse assistant stated the following:

> *We had a list of patients and they had already agreed to answer the screening so we would schedule to do the screening on a specific day when they come to take their medication for tuberculosis.* [Nurse assistant, tuberculosis service]

In addition, as the intervention was tested during PHCP's routine activities, they were unable to implement the screening in all of their shifts because of other responsibilities, rotation to other services, holidays, and community activities. Furthermore, the way the screening task was organized was highly variable between and within the health services. Indeed, there was a high variability regarding the moment in which the screening was performed throughout the consultation, even among PHCPs within the same service. Although some reported doing it at the beginning, others included it at the end of the consultation. Despite this variability, most of them agreed that they would most likely do the screening if their workload was lower or if they had a colleague to relieve the workload during the consultation.

The main challenge was related to specific questions of the screening test. The questions about psychosis were difficult for the PHCPs to score because some patients did not understand the questions about delusional beliefs (see Multimedia Appendix 1, items 19 and 20). The questions about physical symptoms were challenging because of the difficulty of discerning if the symptoms were caused by their physical condition or an underlying mental problem. However, the PHCPs were able to overcome these challenges by using strategies learned in the training sessions, such as explaining the questions to their patients, paraphrasing, and adding a follow-up question for clarity.

Regarding technology, despite PHCPs having different knowledge of tablet-use, no difficulty of use was reported. Additionally, they noted that, compared with a printed version, the app facilitates the task, as they are overburdened with paperwork and believe a new paper-form would add to their current burden.

Overall, when asked about the screening task, the PHCPs gave very positive feedback. Most considered that they were able to provide a more comprehensive care by addressing problems that they may have already identified but did not know how to handle. In addition to this, they mentioned being able to bond with their patients through listening and learning more about them. Some PHCPs also reported being more aware of the importance of their patients' mental health because of its effect on their physical health and personal lives. Furthermore, some of the PHCPs declared that even after ending the intervention, they continued to ask their patients about their emotional well-being. One nurse stated the following:

> *We have been able to improve our consultations and to make patients feel that we want to do more, that is not only a check-up, give them pills and that is all. We have let them know that we care and we can give them more, and it is our intention to do so, even though sometimes we have limitations.* [Nurse, chronic disease service]

Delivery of Results and Referral

Messages conveyed as part of the delivery of the screening results varied widely among PHCPs; nonetheless most of them revolved around the importance of seeking specialized care.

Of the 22 PHCPs, 17 (81%) reported no difficulties with the delivery of screening results. However, some reported difficulties regarding their ability to address mental health issues. Some of the PHCPs found the suicide risk cases particularly challenging and felt unsure about how to handle them. Other less frequent challenges were feeling emotionally affected by the patients' stories and not feeling prepared to offer support, as illustrated in the following quote:

> *For me it was a little difficult to manage the patient that had suicide risk. It was a surprise to know that a patient had that problem, because apparently, she was a strong woman [...]. The patient had this problem during all the pregnancy, and I did not realize it until the screening and that made me feel bad. Maybe more patients come with the same condition and one does not realize it.* [Midwife, antenatal care service]

The PHCPs stated no difficulties with the referral process, with most of the patients being open to seeking specialized care. The PHCPs reported being able to report the actions taken after a positive screening through the app. Furthermore, some PHCPs reported referring patients who, despite having negative screening result, showed symptoms that they deemed relevant.

The referral procedure was different across services. The most common ways were personally escorting the patient to the psychology service, arranging an appointment for the patient, and giving the patient a referral paper for them to seek an appointment with the psychologist. In the services with less patients, such as the tuberculosis and HIV or AIDS, it was more common for the PHCPs to escort their patients to the psychologist's office, whereas it was more common to offer written referral papers in antenatal care and chronic disease services, where only severe cases such as suicide risk were escorted as required by the study protocol.

Though PHCPs were also able to refer their patients to a general practitioner, most of them preferred not to do so, with only five PHCPs reporting having referred a patient to a general practitioner. This was because of the perception that the general practitioner was more suited to treat physical conditions, whereas the psychologist was more appropriate to treat mental health.

Feasibility of Integrating a Mental Health Screening

The majority of PHCPs reported willingness to continue implementing the screening in their services. However, they also noted that time constraints would be a key barrier to do so on a regular basis. Moreover, screening activities could not be reported in the medical records, which meant that these would not count toward their productivity assessments. Most of the PHCPs agreed that they would most likely do the screening when less burdened by workload or if they had a colleague to offer support during their working hours. One nurse stated the following:

> *We are pressured to complete four consultations per hour, and you may think that is not much, but attending four patients take a lot of your time. If the screening is implemented permanently, I think we would have to reduce the number of consultations per hour to three, that way we would have more time. Another solution would be that of our six-hour shift, we only have patients scheduled for five hours, and the spare hour could be used to extend the consultation time for the patients, if needed. There are a lot of possible solutions; it is only a matter of disposition to find them.* [Nurse, chronic diseases service]

Another barrier is the perception by PHCPs that there are not enough psychologists to attend all patients requiring help and that services hours are limited—those who are available may not be present in their services all day because they conduct activities other than consultations and some services only operate in the mornings, thus limiting availability.

To tackle these barriers, PHCPs proposed 3 main suggestions to improve the feasibility and sustainability of the screening: (1) 8 PHCPs (36.4%, 8/22) suggested including the screening as formal procedure in their services with allocated time, thus improving the conditions to accommodate this new task without being overburdened, that is, raise the time limit for each consultation and lower the daily patient workload; (2) 8 PHCPs suggested changes or improvements to the screening test to simplify the questions and make them more comprehensible to the patients; and (3) 3 PHCPs (14.3%, 3/22) advised securing enough specialists to look after the referred patients.

Discussion

Principal Findings

The *Allillanchu Project* aimed to develop and test a multicomponent intervention to improve early detection, opportune referral, and access to treatment of patients with mental disorders attending public PHC services in low-income areas in Lima, Peru. With more than 750 screenings completed in real-world circumstances over a 9-week period, the *Allillanchu Project* showcases the feasibility to integrate the mental health screening into primary care services as a routine procedure. The use of brief screening tools is considered an important first step to integrate mental health care into existing PHC services [41] and is aligned with the proposal of task shifting some mental health care activities, which has been successfully tested in other LMIC settings [42].

With basic training, supervision, and an easy-to-use screening app, PHCPs could be both willing and able, during their regular consultations, to identify symptoms of mental disorders among their patients and refer them to specialized care. The training was key to sensitize PHCPs about their patients' mental health and strengthen their abilities to perform the screening, which is reflected in the improvement of scores at pre- and posttest. Likewise, the supervision was important to provide support when needed, solving doubts with the screening, and the use of technology. Interestingly, the screening was not only effective in many services but also well received by patients, who accepted to be screened and, furthermore, the large majority sought mental health care following the recommendation implemented as part of the project. The results of the screening and health seeking are aligned with the available evidence, which affirms that short-term training with ongoing monitoring and supervision of nonspecialist health workers can improve "confidence, detection, treatment, as well as treatment adherence of individuals with mental disorders" [42]. A relevant component of the intervention was the screening app, which proved to be a successful tool to optimize the detection and referral processes, as well as to monitor its implementation. The benefits of technology to simplify and shorten procedures within the health system have been extensively reported [26-32].

However, it is also true that the inclusion of the screening into the PHCP's routines had both qualitative and quantitative differences. Qualitatively, each service organized the implementation of the screening in different ways, which can be considered a positive sign of adaptation and flexibility to achieve the task [43,44]. Quantitatively, some health services screened many more patients than others. One reason for this is that some services, for example, tuberculosis and HIV or AIDS, have fewer patients than other services. The most significant factor to explain the differing performance among PHCPs was the limited allocated time per consultation combined with the overwhelming patient workload and paperwork burdening only some services, such as antenatal care, leading to saturation, which has been seen in PHC settings in developed countries [45]. These were the two main implementation barriers to regular screening and were especially critical in the antenatal care services. Typically, each service has an established time to complete a set of standard procedures and register them in the clinical records. Therefore, implementing a new task would involve doing more work within the same time period. This observation of a decline in the number of screening in those weeks may be related to the fact that introducing screenings is an activity that competes with many other tasks. Although the general trend suggests that the screening was indeed adopted as part of their routine practices, it also shows that such tasks

were not constant over time, thus calling for ongoing monitoring of these efforts in the future.

The accuracy of the screening was challenged by the ambiguity of some items of the psychosis subsection of the SRQ, the 28-item version recommended by the Ministry of Health. Despite its evidenced ease of use, patients found some of the items difficult to understand and thus, complicated PHCP's scoring when administering the test. This difficulty to understand the questions probably explains the unexpected number of patients testing *positive* for psychotic symptoms, by answering *yes* to any of the four psychosis questions (eg, *Do you feel you can do things that others cannot do, or that you are a particularly important person?*), which indicates a positive case. This highlights the importance of improving the cultural validation of screening tools [41], as well as the importance of continuous training to improve the abilities of PHCPs to formulate the questions and to assess whether the answer provided by the patients matches the symptom described in the question.

In this project, the implementation of the screening revealed that an important proportion of patients regularly attending PHC services to treat a physical condition concomitantly have psychological symptoms that are probably ignored during their consultations. This prevalence, combined with a regular attendance of patients and a closer patient-provider relationship, confirms that performing the screening in PHC is an opportunity for an early detection and opportune referral of patients in need of care [41] An example of the potential benefits of a good screening in primary care is provided by a review of 40 international studies that show that, on average, 3 of 4 suicide victims had contact with a PHCP within the year of suicide, and 45% had it the month before [46].

Existing guidelines recommend screening the general adult population for depression [47,48]. One main goal of our study, besides the introduction of a screening into PHCP's routines, was to increase access to mental health treatment. Remarkably, the referral process had better results than expected. Out of 10 patients, 7 actively sought specialized care, and 5 obtained a consultation with a mental health specialist, thus showing promising results for this type of intervention. Our results contrast sharply with existing practices in Lima, where only 32.8% of those who need mental health treatment report accessing it [49], and with studies in African and European hospitals, where only between 2% and 20% of users with mental disorders actually accessed a specialized consultation [50-52]. Two elements explain this success. First, PHCPs provided opportune advice to positive cases and made a substantial effort to guarantee specialized consultations on the same day of the screening for some of their patients. Second, SMS text messages were highly valued by participants who received them before seeking specialized care, translating into an inexpensive way to reaffirm the advice given by the PHCP.

Yet, several health system barriers remain in place: insufficient information about mental health services and costs, limited time to have an appointment, and scarce availability of mental health specialists, among others, which align with local studies' findings [53-58], as well as our own during the formative study.

Implications for Public Health

One of the most significant strengths of the *Allillanchu Project* was its development and testing through a rigorous multiphase design, set within the current organization of the Peruvian health system, and aligned with the national mental health reform [59]. This is positive in terms of sustainability of the innovation, allowing us to propose context-specific solutions to real-world conditions.

Shifting the detection and referral of people suffering mental disorders to nonspecialized health providers seems to be the right path to alleviate an overburdened health system with scarce specialized resources. However, some conditions need to be improved to make it feasible and sustainable. First, the detection and referral have to be recognized by the public health system as part of the PHCP's tasks, allocating and protecting time for this activity within regular consultations and including it in the patient's clinical record. Second, to ensure the access of referred patients to appropriate treatment, key barriers need to be addressed: informing patients about the availability, time, and cost of services; expanding the hours of consultation; offering free essential mental health care; transferring evidence-based treatments to nonspecialized PHCPs; and increasing the number of mental health specialists. Some of these improvements are already set as goals of Peru's mental health reform [35,59]. Third, it is essential to engage policy makers, health providers, mental health specialists, and users in the effort of integrating mental health care as part of the regular care, caring for potential concerns, and resistance to the innovations by each of these groups.

It is also important to adapt the detection and referral processes to the working dynamics of each service, which vary according to resources and workload. This adaptation may include, for example, for each service to define screening frequency or the prioritization of screening for a subgroup of patients at higher risk of mental disorders.

Training and supervision are both essential to sustain a task-shifting strategy. On the basis of the *Allillanchu Project* experience, training in mental disorders detection and management should be regularly provided to PHCPs, and it should focus on developing skills to manage high-risk patients, to guarantee a more accurate application of the screening tool, and to plan how to introduce the screening with each PHCP in his or her health service. Additionally, as some PHCPs reported being concerned about the impact their patients' accounts had on them, it would be important to include self-care strategies in the training sessions. Considering the time constraints faced by PHCPs and the challenges with the SRQ's psychosis items, it could be appropriate to consider using a shorter instrument focused on fewer disorders such as depression and anxiety because of their higher prevalence among the Peruvian population [49,54-57]. Additionally, exploring alternatives to tackle this issue in future studies, such as self-administered screening test for patients, may contribute to overcome these hurdles.

In terms of sustainability of the mHealth-based screening within the public health system, considering its potential to improve the health system's efficiency, as well as its potential to

centralize patients' information and articulate it with other technology-based health initiatives, for example, electronic health records, it would be advisable to begin to use the mHealth-based screening at a larger scale.

Study's Strengths and Limitations

A major strength of our study was its design and implementation in real-world settings, through accommodating policy makers' preferences, adapting to PHCPs and health services' structures and organization, and creating links between patients who needed specialized care with available care. This strength was paired with the quasi-experimental study design of the implementation phase, which, through using a mixed-methods approach, revealed that no other external factors that could have simultaneously influenced the screening, referral, and access to treatment of participants at risk of developing mental disorders were in place.

Yet, some limitations are worth noticing. For example, the help-seeking responses and access to specialized mental health care, after screening, were self-reported by the patients during a follow-up interview. As we did not have access to the health system records, self-reported information was the best way to collect such information. Another limitation was that only a proportion of participants with a positive screening were eligible to report on perceived effects of the SMS text messages on their help-seeking behavior, largely because of many of them seeking and receiving care before the SMS text messages reached them. This unexpected "positive" result, though less beneficial to the assessment of the SMS text message component in terms of numbers of interviewees, reveals a very promising direction toward the efficacy of screening and referral within nonspecialized services.

Conclusions

The use of a screening app by nonspecialized PHCPs, supported by basic training and supervision, is a feasible procedure and confirms a high prevalence of undiagnosed psychological symptoms among regular users of PHC services. However, for it to be made a routine, the health system needs to formally accommodate it as a PHCP's task and remove its major barriers, particularly time constraints and availability of specialized mental health personnel. To increase their usefulness and sustainability, the detection and referral of cases should be tailored to the workload, resources, and organization of each service. The early detection of psychological symptoms within a regular consultation, followed by adequate advice and motivation, can lead to an important proportion of patients to seek and access specialized care, thus optimizing the use of existing resources and reducing the treatment gap of mental disorders.

Acknowledgments

The authors would like to acknowledge the participation of patients, PHCPs, and mental health specialists in the different stages of the *Allillanchu Project*, as well as the support of the authorities from the Ministry of Health and Social Security System during the approval and implementation of the study. They would also like to extend their appreciation and thanks to various people and institutions for their support: Gabriela Ramos, Gustavo Flórez, and Ricardo Galvez (data collection); Lupe Jara (design and facilitation of training); Jorge Estrada, Oscar Giraldo, and José Olidén (design and development of mHealth components); and Socios en Salud (providing an office during the implementation phase).

This study was supported by the Grand Challenges Canada Global Mental Health initiative under the grant agreement GMH 0335-04. JJM acknowledges receiving additional support from the Alliance for Health Policy and Systems Research (HQHSR1206660), DFID/MRC/Wellcome Global Health Trials (MR/M007405/1), Fogarty International Center (R21TW009982, D71TW010877), International Development Research Center Canada (106887, 108167), Inter-American Institute for Global Change Research (IAI CRN3036), Medical Research Council (MR/P008984/1, MR/P024408/1, MR/P02386X/1), National Cancer Institute (1P20CA217231), National Heart, Lung and Blood Institute (HHSN268200900033C, 5U01HL114180, 1UM1HL134590), National Institute of Mental Health (1U19MH098780), Swiss National Science Foundation (40P740-160366), and the World Diabetes Foundation (WDF15-1224).

Authors' Contributions

FDC, JJM, and RA conceived and designed the study. FDC, AI, MT, SPL, VC, and JJM conducted the study. MT, SPL, VC, and FDC analyzed the data. MT, SPL, VC, FDC, JJM, RA, and AI wrote the manuscript. All authors gave final approval of the version submitted for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Self-Reporting Questionnaire (SRQ).

[PDF File (Adobe PDF File), 30KB - jmir_v20i3e100_app1.pdf]

Multimedia Appendix 2

Follow-up interview guide with patients.

[PDF File (Adobe PDF File), 48KB - jmir_v20i3e100_app2.pdf]

Multimedia Appendix 3

Midterm and postintervention interview guide with PHCP (primary health care provider).

[PDF File (Adobe PDF File), 78KB - jmir_v20i3e100_app3.pdf]

Multimedia Appendix 4

Analysis of patients' and PHCPs' (primary health care provider) interviews.

[PDF File (Adobe PDF File), 48KB - jmir_v20i3e100_app4.pdf]

References

1. World Health Organization. Global Status Report on Non-Communicable Diseases 2010. Geneva: WHO; 2011.
2. Vega P, Sweetland A, Acha J, Castillo H, Guerra D, Smith FMC, et al. Psychiatric issues in the management of patients with multidrug-resistant tuberculosis. Int J Tuberc Lung Dis 2004 Jun;8(6):749-759. [Medline: 15182146]
3. Bao-Alonso M, Vega-Dienstmaier J, Saona-Ugarte P. Prevalencia de depresión durante la gestación. Revista de Neuro-Psiquiatría 2010;73(3):95-103. [doi: 10.20453/rnp.v73i3.1710]
4. Cripe SM, Sanchez S, Lam N, Sanchez E, Ojeda N, Tacuri S, et al. Depressive symptoms and migraine comorbidity among pregnant Peruvian women. J Affect Disord 2010 Apr;122(1-2):149-153 [FREE Full text] [doi: 10.1016/j.jad.2009.07.014] [Medline: 19695709]
5. Luna Matos ML, Salinas Piélago J, Luna Figueroa A. [Major depression in pregnant women served by the National Materno-Perinatal Institute in Lima, Peru]. Rev Panam Salud Publica 2009 Oct;26(4):310-314 [FREE Full text] [Medline: 20107678]
6. Wu DY, Munoz M, Espiritu B, Zeladita J, Sanchez E, Callacna M, et al. Burden of depression among impoverished HIV-positive women in Peru. J Acquir Immune Defic Syndr 2008 Aug 01;48(4):500-504. [doi: 10.1097/QAI.0b013e31817dc3e9] [Medline: 18614919]
7. Constantino-Cerna A, Bocanegra- Malca M, León-Jiménez F, Díaz-Vélez C. Frecuencia de depresión y ansiedad en pacientes con diabetes tipo 2 atendidos en un hospital general de Chiclayo. Rev Med Hered 2014 Dec 18;25(4):196-203. [doi: 10.20453/rmh.v25i4.2176]
8. Urrutia-Aliano D, Segura ER. [Depressive symptoms and type 2 diabetes mellitus in outpatients of an Armed Forces hospital in Lima, Peru, 2012: a cross-sectional study]. Medwave 2016 Apr 29;16(3):e6435. [Medline: 27144926]
9. Crispín-Trebejo B, Robles-Cuadros MC, Bernabé-Ortiz A. Association between depression and glycemic control among type 2 diabetes patients in Lima, Peru. Asia Pac Psychiatry 2015 Dec;7(4):419-426 [FREE Full text] [doi: 10.1111/appy.12190] [Medline: 26037488]
10. Rayner L, Price A, Evans A, Valsraj K, Higginson IJ, Hotopf M. Antidepressants for depression in physically ill people. Cochrane Database Syst Rev 2010(3):CD007503. [doi: 10.1002/14651858.CD007503.pub2] [Medline: 20238354]
11. Stein MB, Cox BJ, Afifi TO, Belik S, Sareen J. Does co-morbid depressive illness magnify the impact of chronic physical illness? A population-based perspective. Psychol Med 2006 May;36(5):587-596. [doi: 10.1017/S0033291706007239] [Medline: 16608557]
12. Unützer J, Harbin H, Schoenbaum M, Druss B. CHCS. United States: Medicaid; 2013 May. The Collaborative Care Model: An Approach for Integrating Physical and Mental Health Care in Medicaid Health Homes URL: https://www.chcs.org/media/HH_IRC_Collaborative_Care_Model__052113_2.pdf [accessed 2018-02-27] [WebCite Cache ID 6xXaVem22]
13. Evans EC, Bullock LFC. Optimism and other psychosocial influences on antenatal depression: a systematic review. Nurs Health Sci 2012 Sep;14(3):352-361. [doi: 10.1111/j.1442-2018.2012.00700.x] [Medline: 22762538]
14. Field T. Prenatal depression effects on early development: a review. Infant Behav Dev 2011 Feb;34(1):1-14. [doi: 10.1016/j.infbeh.2010.09.008] [Medline: 20970195]
15. Freed RD, Chan PT, Boger KD, Tompson MC. Enhancing maternal depression recognition in health care settings: a review of strategies to improve detection, reduce barriers, and reach mothers in need. Fam Syst Health 2012 Mar;30(1):1-18. [doi: 10.1037/a0027602] [Medline: 22429076]
16. Marcus SM. Depression during pregnancy: rates, risks and consequences--Motherisk Update 2008. Can J Clin Pharmacol 2009;16(1):e15-e22. [Medline: 19164843]
17. Castillo H. Situación de la atención de salud mental en el Perú: Resultados preliminares del Estudio Epidemiológico de Salud Mental en Lima Replicación 2012. 2012 Presented at: Ciclo de Conferencias de la Academia Nacional de Medicina y el Instituto Nacional de Salud; 2012; Lima, Peru.

18. Velásquez A. The burden of disease and injuries in Peru and the priorities of the universal assurance essential plan. Rev Peru Med Exp Salud Publica 2009 Jan;26(2):222-231 [FREE Full text]
19. Diez-Canseco F, Ipince A, Toyama M, Benate-Galvez Y, Galán-Rodas E, Medina-Verástegui JC, et al. [Integration of mental health and chronic non-communicable diseases in Peru: challenges and opportunities for primary care settings]. Rev Peru Med Exp Salud Publica 2014;31(1):131-136 [FREE Full text] [Medline: 24718538]
20. World Health Organization. Integrating Mental Health into Primary Care: A Global Perspective. Geneva: WHO; 2008:1-206.
21. Buttorff C, Hock RS, Weiss HA, Naik S, Araya R, Kirkwood BR, et al. Economic evaluation of a task-shifting intervention for common mental disorders in India. Bull World Health Organ 2012 Nov 01;90(11):813-821 [FREE Full text] [doi: 10.2471/BLT.12.104133] [Medline: 23226893]
22. Vital Wave Consulting. mHealth for Development: The Opportunity of Mobile Technology for Healthcare in the Developing World. Washington, D. C. Washington, D.C. and Berkshire, UK: UN Foundation-Vodafone Foundation Partnership; 2009.
23. Beratarrechea A, Diez-Canseco F, Irazola V, Miranda J, Ramirez-Zea M, Rubinstein A. Use of m-Health Technology for Preventive Interventions to Tackle Cardiometabolic Conditions and Other Non-Communicable Diseases in Latin America-Challenges and Opportunities. Prog Cardiovasc Dis 2016;58(6):661-673. [doi: 10.1016/j.pcad.2016.03.003] [Medline: 27041078]
24. Maiorana A, Kegeles S, Salazar X, Konda K, Silva-Santisteban A, Cáceres C. 'Proyecto Orgullo', an HIV prevention, empowerment and community mobilisation intervention for gay men and transgender women in Callao/Lima, Peru. Glob Public Health 2016;11(7-8):1076-1092 [FREE Full text] [doi: 10.1080/17441692.2016.1161814] [Medline: 27373578]
25. Torres LD, Barrera AZ, Delucchi K, Penilla C, Pérez-Stable EJ, Muñoz RF. Quitting smoking does not increase the risk of major depressive episodes among users of Internet smoking cessation interventions. Psychol Med 2010 Mar;40(3):441-449 [FREE Full text] [doi: 10.1017/S0033291709990560] [Medline: 19627638]
26. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
27. 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]
28. Griffiths KM, Farrer L, Christensen H. The efficacy of internet interventions for depression and anxiety disorders: a review of randomised controlled trials. Med J Aust 2010 Jun 7;192(11 Suppl):S4-11. [Medline: 20528707]
29. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 2013 Jan;10(1):e1001363 [FREE Full text] [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
30. Adepoju IO, Albersen BJ, De Brouwere V, van Roosmalen J, Zweekhorst M. mHealth for Clinical Decision-Making in Sub-Saharan Africa: A Scoping Review. JMIR Mhealth Uhealth 2017 Mar 23;5(3):e38 [FREE Full text] [doi: 10.2196/mhealth.7185] [Medline: 28336504]
31. Agarwal S, Perry HB, Long L, Labrique AB. Evidence on feasibility and effective use of mHealth strategies by frontline health workers in developing countries: systematic review. Trop Med Int Health 2015 Aug;20(8):1003-1014 [FREE Full text] [doi: 10.1111/tmi.12525] [Medline: 25881735]
32. Vedanthan R, Blank E, Tuikong N, Kamano J, Misoi L, Tulienge D, et al. Usability and feasibility of a tablet-based Decision-Support and Integrated Record-keeping (DESIRE) tool in the nurse management of hypertension in rural western Kenya. Int J Med Inform 2015 Mar;84(3):207-219 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.12.005] [Medline: 25612791]
33. Beratarrechea A, Lee AG, Willner JM, Jahangir E, Ciapponi A, Rubinstein A. The impact of mobile health interventions on chronic disease outcomes in developing countries: a systematic review. Telemed J E Health 2014 Jan;20(1):75-82 [FREE Full text] [doi: 10.1089/tmj.2012.0328] [Medline: 24205809]
34. Liew S, Tong SF, Lee VKM, Ng CJ, Leong KC, Teng CL. Text messaging reminders to reduce non-attendance in chronic disease follow-up: a clinical trial. Br J Gen Pract 2009 Dec;59(569):916-920 [FREE Full text] [doi: 10.3399/bjgp09X472250] [Medline: 19712544]
35. Miranda JJ, Diez-Canseco F, Araya R, Cutipe Y, Castillo H, Herrera V, et al. Transitioning mental health into primary care. Lancet Psychiatry 2017 Feb;4(2):90-92. [doi: 10.1016/S2215-0366(16)30350-9] [Medline: 28137383]
36. Creswell JW, Plano Clark VL. Designing and Conducting Mixed Methods Research. California: SAGE; 2007.
37. Véliz J. Validez y Confiabilidad del Cuestionario de Auto-reportaje de Síntomas (S.R.Q.) en personal de Sanidad de la Policía Nacional. Lima, Peru: Universidad Peruana Cayetano Heredia; 1996.
38. Toyama M, Diez-Canseco F, Busse P, Del Mastro I, Miranda JJ. Design and content validation of a set of SMS to promote seeking of specialized mental health care within the Allillanchu Project. Glob Health Epidemiol Genom 2018 Jan 31;3:e2. [doi: 10.1017/gheg.2017.18]
39. World Health Organization. mhGAP Intervention Guide for Mental, Neurological and Substance Use Disorders in Non-Specialized Health Settings. Geneva: WHO; 2011.

40. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Med Res Methodol 2013 Sep 18;13:117 [FREE Full text] [doi: 10.1186/1471-2288-13-117] [Medline: 24047204]
41. Ali G, Ryan G, De Silva MJ. Validated Screening Tools for Common Mental Disorders in Low and Middle Income Countries: A Systematic Review. PLoS One 2016;11(6):e0156939 [FREE Full text] [doi: 10.1371/journal.pone.0156939] [Medline: 27310297]
42. Kakuma R, Minas H, van Ginneken N, Dal Poz MR, Desiraju K, Morris JE, et al. Human resources for mental health care: current situation and strategies for action. Lancet 2011 Nov 05;378(9803):1654-1663. [doi: 10.1016/S0140-6736(11)61093-3] [Medline: 22008420]
43. Barry MM, Clarke AM, Petersen I. Promotion of mental health and prevention of mental disorders: priorities for implementation. East Mediterr Health J 2015 Sep 28;21(7):503-511. [Medline: 26442891]
44. Zechmeister I, Kilian R, McDaid D, MHEEN group. Is it worth investing in mental health promotion and prevention of mental illness? A systematic review of the evidence from economic evaluations. BMC Public Health 2008 Jan 22;8:20 [FREE Full text] [doi: 10.1186/1471-2458-8-20] [Medline: 18211677]
45. Hobbs FDR, Bankhead C, Mukhtar T, Stevens S, Perera-Salazar R, Holt T, National Institute for Health Research School for Primary Care Research. Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007-14. Lancet 2016 Jun 04;387(10035):2323-2330 [FREE Full text] [doi: 10.1016/S0140-6736(16)00620-6] [Medline: 27059888]
46. Luoma JB, Martin CE, Pearson JL. Contact with mental health and primary care providers before suicide: a review of the evidence. Am J Psychiatry 2002 Jun;159(6):909-916 [FREE Full text] [doi: 10.1176/appi.ajp.159.6.909] [Medline: 12042175]
47. Siu AL, US Preventive Services Task Force (USPSTF), Bibbins-Domingo K, Grossman DC, Baumann LC, Davidson KW, et al. Screening for Depression in Adults: US Preventive Services Task Force Recommendation Statement. JAMA 2016 Jan 26;315(4):380-387. [doi: 10.1001/jama.2015.18392] [Medline: 26813211]
48. The Community Preventive Services Task Force. The Community Guide. 2010. Mental Health and Mental Illness: Collaborative Care for the Management of Depressive Disorders URL: https://www.thecommunityguide.org/findings/mental-health-and-mental-illness-collaborative-care-management-depressive-disorders [accessed 2018-02-20] [WebCite Cache ID 6xNfmOsUW]
49. Instituto Nacional de Salud Mental. 2013. Estudio Epidemiológico de Salud Mental en Lima Metropolitana y Callao URL: http://www.insm.gob.pe/investigacion/archivos/estudios/2012%20ASM%20-EESM%20-LM.pdf [accessed 2018-03-05] [WebCite Cache ID 6xggb2JNi]
50. Getinet W, Dessalegn Y. Prevalence and associated factors of common mental disorders among patients admitted in Gondar university hospital medical and surgical wards, northwest Ethiopia, 2013. AJPN 2014;2(3):43-49.
51. Hansen MS, Fink P, Frydenberg M, Oxhøj M, Søndergaard L, Munk-Jørgensen P. Mental disorders among internal medical inpatients: prevalence, detection, and treatment status. J Psychosom Res 2001 Apr;50(4):199-204. [Medline: 11369025]
52. Mustafa MA, Fawzy AR. Psychiatric morbidity and consultations among medical and surgical inpatients in a general hospital in Kuwait. The Arab Journal of Psychiatry 2009;20(2):87-108.
53. Instituto Nacional de Salud Mental. 2003. Estudio Epidemiológico de Salud Mental en la Sierra Peruana 2003 URL: http://www.insm.gob.pe/investigacion/archivos/estudios/2003-ASM-EESM-SP/files/res/downloads/book.pdf [accessed 2018-03-05] [WebCite Cache ID 6xggkguzn]
54. Instituto Nacional de Salud Mental. 2004. Estudio Epidemiológico de Salud Mental en Ayacucho 2003 URL: http://www.insm.gob.pe/investigacion/archivos/estudios/2004-ASM-EESM-I/files/res/downloads/book.pdf [accessed 2018-03-05] [WebCite Cache ID 6xggx2W78]
55. Instituto Nacional de Salud Mental. Estudio Epidemiológico de Salud Mental en la Selva Peruana. Anales de Salud mental 2005;21(1-2):1-214.
56. Instituto Nacional de Salud Mental. Estudio Epidemiológico de Salud Mental en Fronteras 2005. Anales de Salud Mental 2006;22(1-2):1-227.
57. Instituto Nacional de Salud Mental. Estudio Epidemiológico de Salud Mental en Lima Rural 2007. Anales de Salud Mental 2007;24(1-2):1-247.
58. Instituto Nacional de Salud Mental. Estudio Epidemiológico de Salud Mental en la Costa Peruana 2006. Anales de Salud Mental 2007;23(1-2):1-226.
59. Toyama M, Castillo H, Galea JT, Brandt LR, Mendoza M, Herrera V, et al. Peruvian Mental Health Reform: A Framework for Scaling-up Mental Health Services. Int J Health Policy Manag 2017 Jan 22;6(9):501-508 [FREE Full text] [doi: 10.15171/ijhpm.2017.07] [Medline: 28949462]

Abbreviations

LMIC: low- and middle-income countries
mHealth: mobile health
PHC: primary health care

PHCP: primary health care provider
SMS: short message service
SRQ: Self Report Questionnaire

Edited by G Eysenbach; submitted 06.11.17; peer-reviewed by J Galea, R Kanter, TR Soron; comments to author 13.12.17; accepted 02.01.18; published 15.03.18.

Please cite as:
Diez-Canseco F, Toyama M, Ipince A, Perez-Leon S, Cavero V, Araya R, Miranda JJ
Integration of a Technology-Based Mental Health Screening Program Into Routine Practices of Primary Health Care Services in Peru (The Allillanchu Project): Development and Implementation
J Med Internet Res 2018;20(3):e100
URL: http://www.jmir.org/2018/3/e100/
doi:10.2196/jmir.9208
PMID:29588272

Original Paper

Evaluating Patient Perspectives of Provider Professionalism on Twitter in an Academic Obstetrics and Gynecology Clinic: Patient Survey

Rosalyn E Maben-Feaster[1], MD, MPH; R Brent Stansfield[2], PhD; AnneMarie Opipari[3]; Maya M Hammoud[1], MD

[1]Department of Obstetrics and Gynecology, University of Michigan, Ann Arbor, MI, United States

[2]Graduate Medical Education, Wayne State University School of Medicine, Detroit, MI, United States

[3]University of Michigan Medical School, Ann Arbor, MI, United States

Corresponding Author:
Rosalyn E Maben-Feaster, MD, MPH
Department of Obstetrics and Gynecology
University of Michigan
L4000 University Hospital South
1500 E Medical Center Dr
Ann Arbor, MI, 48109
United States
Phone: 1 734 936 3110
Fax: 1 734 763 5992
Email: rmaben@med.umich.edu

Abstract

Background: One-third of Americans use social media websites as a source of health care information. Twitter, a microblogging site that allows users to place 280-character posts—or tweets—on the Web, is emerging as an important social media platform for health care. However, most guidelines on medical professionalism on social media are based on expert opinion.

Objective: This study sought to examine if provider Twitter profiles with educational tweets were viewed as more professional than profiles with personal tweets or a mixture of the two, and to determine the impact of provider gender on perceptions of professionalism in an academic obstetrics and gynecology clinic.

Methods: This study randomized obstetrics and gynecology patients at the University of Michigan Von Voigtlander Clinic to view one of six medical provider Twitter profiles, which differed in provider gender and the nature of tweets. Each participant answered 10 questions about their perception of the provider's professionalism based on the Twitter profile content.

Results: The provider profiles with educational tweets alone received higher mean professionalism scores than profiles with personal tweets. Specifically, the female and male provider profiles with exclusively educational tweets had the highest and second highest overall mean professionalism ratings at 4.24 and 3.85, respectively. In addition, the female provider profiles received higher mean professionalism ratings than male provider profiles with the same content. The female profile with mixed content received a mean professionalism rating of 3.38 compared to 3.24 for the male mixed-content profile, and the female profile with only personal content received a mean professionalism rating of 3.68 compared to 2.68 for the exclusively personal male provider profile.

Conclusions: This study showed that in our obstetrics and gynecology clinic, patients perceived providers with educational profiles as more professional than those with a mixture of educational and personal tweets or only personal tweets. It also showed that our patient population perceived the female provider with educational tweets to be the most professional. This study will help inform the development of evidence-based guidelines for social media use in medicine as it adds to the growing body of literature examining professionalism and social media.

(J Med Internet Res 2018;20(3):e78) doi:10.2196/jmir.8056

KEYWORDS

patients; social networking sites; professionalism; surveys and questionnaires; perception; Twitter

Introduction

Social media is a form of online communication, such as websites for social networking and microblogging, through which users can create online communities to share information, ideas, personal messages, and other content [1,2]. In a 2014 study, the Pew Research Center showed that 74% of Americans on the Internet use social media sites [3]. Furthermore, one-third of Americans use social media for health care discussions, according to a study by PricewaterhouseCoopers [4]. Facebook, a platform for users to share their stories and connect to other people, is the most popular site, with 1.04 billion daily users worldwide [5,6]. Other commonly used tools include YouTube, which is a social media platform that allows users to discover, watch, and share videos, and Twitter, a microblogging site that allows users to place 280-character posts—or tweets—on the Web [7,8]. YouTube has approximately one billion users worldwide, while Twitter has 320 million active users [7,8]. A Health Research Institute consumer survey showed that Facebook and YouTube are the most commonly used social media tools for consumers to view health-related information [4]. However, Twitter is emerging as one of the leading social media platforms for health care. It has shown significant growth, with 460,000 new accounts created on average per day [4]. Social media offers a number of opportunities for health care organizations and health care professionals. Twitter alone has been shown to have more than 140 different uses in health care [9] including disaster alerting and response, diabetes management, and drug safety [10]. On an individual level, 90% of physicians use social media for personal reasons and 67% use it professionally [11,12]. Thus, it serves as a way for health care providers to provide health education, connect with patients, and increase market share via a unique platform that allows information to be disseminated beyond the capabilities of traditional digital media—such as webpages—which makes it an attractive tool for organizations and individual professionals [13]. However, this capacity to reach a large audience also increases the likelihood of unknown users interacting with the site [13].

Many professional organizations such as the American Medical Association, the Federation for State Medical Boards, the American Board of Internal Medicine, the American College of Obstetricians and Gynecologists, and many individual health care organizations have developed guidelines to help health care organizations and providers create social media presences that discourage posting of inaccurate information, avoid damaging professional identities, preserve patient privacy and the provider-patient relationship, and avoid ligation [9]. However, the bulk of these guidelines are created based on expert opinion. Thus, there is a need for research into the evolution of professionalism in the digital era—also known as e-professionalism [14]. There are numerous papers discussing the issue of professionalism in social media. However, there is limited data on how a provider's social media presence impacts a patient's perception of that provider's professionalism. A 2014 study by Jain and colleagues at the University of Michigan evaluated what medical students, doctors, and the public felt was unprofessional for medical students to post on Facebook [15]. The results showed that the public and faculty had lower thresholds for what was considered appropriate and that this was also related to how comfortable they would be with these students caring for them [15]. Interestingly, doctors, females, and older individuals were less permissive regarding the appropriateness of content [15]. Clyde and colleagues examined how a physician's Facebook profile can impact a potential patient's impression of that provider's professionalism [16] and found that personal profiles containing healthy behavior were rated as most professional, followed by profiles with strictly professional content. Unhealthy personal profiles were rated as least professional [16]. In addition, profiles of female providers were rated more professional across all profile types [16].

However, data are lacking on the public's perception of provider professionalism in the context of Twitter. There is also limited research on how obstetrics and gynecology patients and providers use social media and how that use relates to medical professionalism. Our study sought to help close this gap and inform the discussion on how health care providers should use social media. Specifically, we aimed to determine whether a provider Twitter profile with educational tweets was viewed as more professional than a Twitter profile with personal tweets or a mixture of the two, as determined by patients in an academic obstetrics and gynecology clinic. It also sought to determine whether these patients would perceive the Twitter profiles of female providers to be more professional than the Twitter profiles of male providers, regardless of the content of the profile.

Methods

Ethics

This study, including the introduction letter and survey instrument, was reviewed and determined exempt by the Medical School Institutional Review Board at the University of Michigan. A signed consent was not required, but each survey began with a letter explaining that this was a voluntary research study. This letter also explained how the participant's confidentiality would be maintained, that participation would not affect their care, and that the participant could receive the study results at study completion if they wished. Finally, the letter also offered them the opportunity to be entered in a draw for one of five gift certificates valued at US $100. The contact information for those participants that entered the draw was kept separate from the survey to protect their confidentiality. Participants were able to withdraw by their own request at any time by simply not submitting the survey.

Eligibility Criteria and Study Setting

In July 2012, 200 surveys were distributed to women receiving care in the Obstetrics and Gynecology clinic at the University of Michigan Von Voigtlander Women's Hospital in Ann Arbor, Michigan. In order to participate, patients had to be at least 18 years or older and able to read English.

Study Design

Patients were randomized to receive one of six different printed survey packets at the time of appointment check-in. Each survey packet included a letter describing the study, questions regarding

demographics and social media use, a color screenshot of one of six different medical provider Twitter profiles, and a 10-question survey instrument looking at the patient's attitude about the professionalism of the provider whose profile was viewed. Participants submitted the completed packet prior to leaving the clinic.

Medical Provider Twitter Profile Creation

Each of the six profiles was created on Twitter. The profiles all had the same profile picture (stethoscope clipart), background, and contained 8 tweets. However, the profiles differed with regard to profile name and content. Specifically, the names were selected to reflect the profiles of three female physicians and 3 male physicians. Twitter does not allow the creation of multiple profiles with the exact same name, thus the female provider profiles had three permutations of a similar first name but the same last name (Ashley, Ashlee, and Ashleigh Scott, MD). The male profiles followed a similar pattern (Jahn, John, and Jon Scott, MD). For each gender, the Twitter profiles were designed to fit into one of three categories: educational only (Ashlee and John), personal only (Ashley and Jahn), or a 50/50 mixture of the two (Ashleigh and Jon). The tweets used were adapted from tweets posted by self-identified obstetrician gynecologists on Twitter. The 8 tweets selected were the same within the content group regardless of gender. For example, the male provider's educational only Twitter profile had the exact same tweets as the female provider's educational only Twitter profile (see Multimedia Appendix 1).

Rating Scale for Provider Professionalism

Patients were asked to rate 10 statements about each provider's professionalism based on their tweets. The professionalism statements were developed from the American Board of Internal Medicine's 10 professional responsibilities: professional competence, honesty with patients, patient confidentiality, maintaining appropriate relationships with patients, improving quality of care, improving access to care, just distribution of finite resources, scientific knowledge, maintaining trust by managing conflicts of interest, and professional responsibilities [17,18].

A 7-point scale was used to rate the responses to each of the 10 professionalism statements:

1. This provider seems like he/she has the skills to take care of my health care needs (skill).
2. This provider seems like he/she would be honest with me (honesty).
3. This provider seems like he/she would keep my health information private (privacy).
4. This provider seems like he/she would maintain appropriate boundaries with patients and other health care providers (boundary).
5. This provider seems like he/she would work to improve the quality of health care (work to improve).
6. This provider seems like he/she would work to provide good access to health care (access).
7. This provider seems like he/she would use health resources fairly and appropriately (resource use).
8. This provider seems like he/she knows what he/she is doing (competent).
9. This provider seems like he/she knows how to maintain the patient's trust while balancing conflicting interests (trust).
10. This provider seems like he/she follows through on his/her professional responsibilities (responsible).

Statistical Analysis

In the Johnson study, there was a nonparametric correlation between perceived credibility and whether a teacher posted social or scholarly content [19]. Given the use of nonparametric statistics in the Johnson study, the effect size for a power analysis was difficult to estimate, but we conservatively estimated a moderate effect according to Cohen (d=0.3). We calculated a power of 80% with 87 participants [19]. Adding gender and its interaction, a sample size of 100 gave a power of 80% to detect an overall effect size of eta squared=0.105 [19]. We predicted that if there was an effect of professionalism and gender in an interaction, and these together explained 10.5% of the variance, we would find the effect 80 times out of 100 experiments [19]. The data was analyzed using analysis of variance, with a P value of .05 considered to be statistically significant. The data from all participants who submitted answers to the professionalism section of the survey were included in the analysis.

Results

During the 4 weeks of survey collection, 200 surveys were distributed and 134 were returned completed, giving a response rate of 67%. We were unable to collect data about the patients who opted not to complete the survey given how the packets were distributed.

The demographics of the population surveyed reflected our clinic's overall population. In general, our patients are primarily young, college-educated, married, non-Hispanic white women (Table 1). The majority (127/134, 94.8%) of participants were under age 50, three-quarters (99/134, 73.9%) identified as white, two-thirds (87/134, 64.9%) were married, and the majority (120/134, 89.6%) completed at least some college (Table 1). With regard to annual household income, 34.3% (46/134) made more than US $75,000 per year and 34.3% (46/134) made less than US $50,000 per year.

With respect to social media use, 91.0% (122/134) of participants used social media, with Facebook, blogs, and Twitter being the most popular. However, only 20.9% (28/134) of participants used social media for health care purposes. The uses indicated by our participants included gaining knowledge about conditions or treatments, support groups, sharing experiences, and receiving or giving advice.

When examining the ratings of the 10 professionalism statements (ie, skill, honesty, privacy, boundary, work to improve, access, resource use, competent, trust, responsible), we found a statistically significant difference for 6 of the 10 dimensions, including skill, work to improve, access, resource use, competent, and responsible (Table 2).

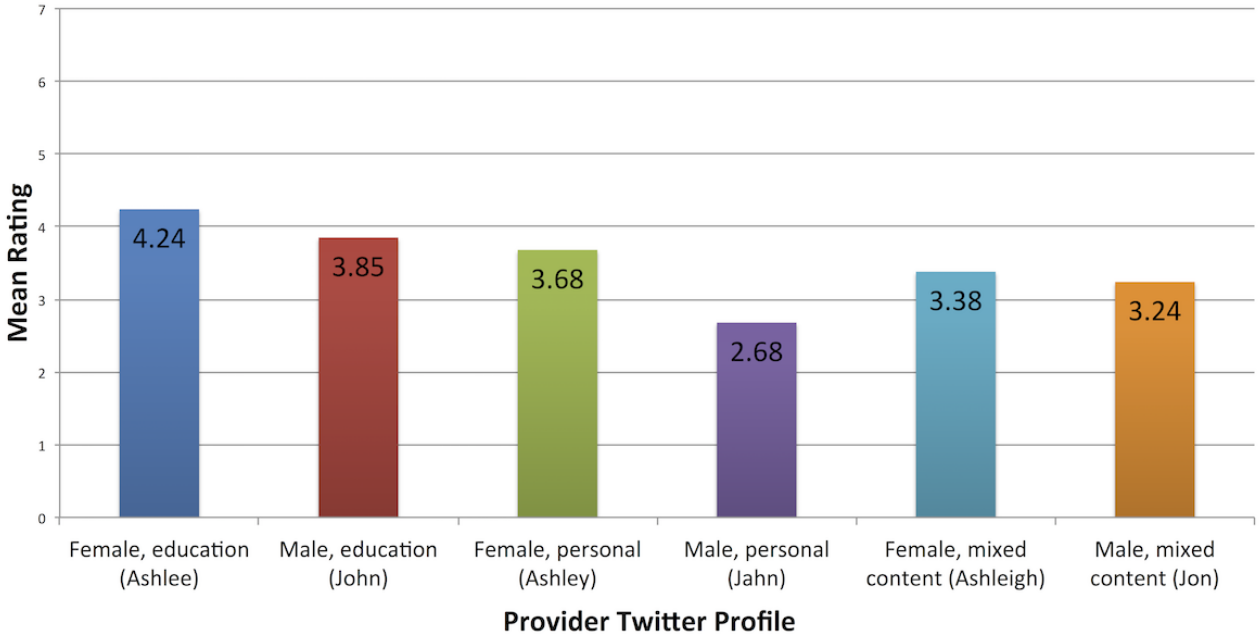
Table 1. Demographics of study participants (N=134).

Characteristics	n (%)
Race/Ethnicity	
White	99 (74)
Black	11 (8)
Asian	13 (10)
American Indian or Alaskan Native	2 (1)
Other	5 (4)
Hispanic or Latino	5 (4)
Marital status	
Married	87 (65)
Single	39 (29)
Separated	2 (1)
Divorced	2 (1)
Widowed	1 (1)
Age	
18-29	52 (39)
30-49	75 (56)
50-64	9 (7)
≥65	3 (2)
Education level	
Not a high school graduate	4 (3)
High school graduate	10 (7)
Some college	120 (90)

Table 2. Survey results: mean ratings for professionalism statements by profile types. Ratings based on 7-point scale.

	Female, education (Ashlee)	Male, education (John)	Female, personal (Ashley)	Male, personal (Jahn)	Female, mixed (Ashleigh)	Male, mixed (Jon)
Skill	4.14	3.65	3.30	2.26	3.18	2.58
Honesty	4.05	3.47	3.96	2.91	3.71	3.53
Privacy	3.90	3.29	3.65	2.78	3.24	3.42
Boundary	4.10	3.65	3.91	2.65	3.06	3.26
Work to improve	4.40	4.59	3.57	2.50	3.65	3.16
Access	4.50	4.59	3.78	2.78	3.76	3.21
Resource use	4.45	3.94	3.78	2.74	3.88	3.28
Competent	4.29	3.82	3.65	2.48	3.29	3.82
Trust	4.15	3.59	3.57	2.78	3.12	3.00
Responsible	4.45	3.88	3.61	2.87	2.94	3.11

Figure 1. Survey results: overall mean ratings by profile type.



In general, profiles with educational tweets received higher professionalism ratings compared to those with personal tweets (Table 2). Also, female providers received higher ratings than male providers with similar profiles (Table 2). Specifically, the study population rated the female provider with educational tweets (Ashlee Scott, MD) as more professional than the male provider with personal tweets (Jahn Scott, MD) and the male provider with a mixture of educational and personal tweets (Jon Scott, MD) in each of these dimensions (Table 2). The female provider with educational tweets was also rated more responsible than the female provider with a mixture of educational and personal tweets (Ashleigh Scott, MD) (Table 2). Rating items were highly intercorrelated, suggesting that the ratings reflect the same construct or a set of highly related constructs. This intercorrelation justified computing a mean rating across items as an estimate of the participant's perception of the physician's overall quality. At 5% significance level, the data provide sufficient evidence to conclude that a difference exists between the mean rating among the six profiles (P=.002). In addition, at a 5% significance level, the data provide sufficient evidence to conclude that mean ratings for the female educational Twitter profile were higher than those for the male personal profile (P=.001) and the male mixed profile (P=.03) (Figure 1). Specifically, the female provider profile with educational tweets had a mean rating of 4.243 compared to 3.847 for the male provider profile with the same tweets (Figure 1). The mean rating for the female provider profile with personal tweets was 3.678 compared to 2.675 for the male provider profile with the same tweets (Figure 1). Finally, the female provider profile with a mixture of educational and personal tweets had a mean rating of 3.383 compared to a mean rating of 3.237 for the male provider profile with the same tweets (Figure 1).

Discussion

Principal Findings

Many physicians use social media for both personal and business uses [11,12]. However, data are limited on how a provider's social media profile impacts the patient's perception of that provider's professionalism. Our study sought to help close this gap in the obstetrics and gynecology patient population. In this study, obstetrics and gynecology patients were randomized to view a screenshot from the Twitter profile of one of six different fictitious providers and then rate their professionalism based on the content of the tweets viewed. In general, profiles with educational tweets received higher professionalism ratings from our study participants than profiles with mixed content or purely personal tweets. Specifically, the female and male provider profiles with exclusively educational tweets had the highest and second highest overall mean professionalism ratings at 4.24 and 3.85 respectively. This is consistent with what would have been expected based on the traditional definition of medical professionalism outside the context of social media. In keeping with this theory, the mean professionalism score among the male provider profiles decreased as the content of the profile became more personal in nature with the exclusively personal Twitter profile having the lowest professionalism score of 2.68. However, among the female provider profiles, a slightly different pattern was seen. The purely educational Twitter profile had the highest professionalism rating. However, the female provider profile with the second highest mean professionalism rating was actually the provider profile with exclusively personal tweets at 3.68 rather than the mixed content Twitter profile at 3.38. This discrepancy may be driven by the fact that the patient population surveyed was all female and these patients may have identified more with the persona of the female provider with personal tweets, resulting in a higher professionalism score. This may also explain why the female provider profiles had

higher mean professionalism scores when compared to the male provider profiles with the same content.

Limitations

There were a number of limitations to this study. First, the study was conducted in an obstetrics and gynecology clinic and all of the study participants identified as female. This may have biased our finding that female providers were viewed as more professional since the patients may have identified more with providers of the same gender. Interestingly, Jain and colleagues studied medical student Facebook profiles and found that female participants tended to be less permissive regarding the appropriateness of profile content [15]. The Facebook-based professionalism study conducted by Clyde and colleagues also found that those surveyed viewed female providers as more professional and their study population included both female and male participants [16]. This suggests that more research is needed to determine why female gender alone creates a greater perception of professionalism. The majority of the participants in this study were under age 50 (95%), which limits the generalizability of these findings to an older population. However, Pew Research Center data suggest that Twitter use is less common in persons over age 50 [3]. Their data show that only 11% of Internet users over age 50 use Twitter, compared to 30% of Internet users under age 50 [3]. This suggests that in older populations, Twitter is less likely to be used for health-related information and provider selection. Finally, at the time of study design and implementation, there was no pre-existing measurement scale available for evaluation of perceptions of professionalism. Therefore, we developed our own survey instrument, which had not been validated in other studies. However, the ratings were highly intercorrelated, suggesting that the ratings reflect the same or a highly similar construct, which supports the instrument's validity. Since conducting our study, another scale to measure perceived professionalism was developed—the First Impressions of Medical Professionalism (FIMP) scale [16]. This creates an opportunity to confirm the findings of our study using a different instrument.

Conclusion

When patients view a provider's profile on a social media platform, they do not necessarily discern between whether the provider's profile is for personal or business use. In addition, it is difficult to control who may be able to view a given profile. Professionalism serves as the foundation of the patient-provider relationship. If it is eroded prior to the patient entering the clinic, due to a provider's social media presence, this can have implications with regard to care. However, social media platforms also allow medical providers to interact and reach patients in a unique way, which may improve care. Given this delicate balance, it is important to expand the body of knowledge on medical professionalism in the context of social media. In an academic obstetrics and gynecology clinic, we found patients identifying as female perceived providers with purely educational Twitter feeds as more professional than those with mixed content or purely personal tweets. To our knowledge, this is the first study to look at the issue of patient-perceived professionalism among providers who tweet and the impact of provider gender on this perception. It provides a foundation for further research into how this technology impacts our ability to educate patients and each other.

Acknowledgments

This research was supported by the University of Michigan 2012 Dr Milton Goldrath Resident Research Support Award.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Medical provider Twitter profile screenshots as viewed by study participants.

[PDF File (Adobe PDF File), 1MB - jmir_v20i3e78_app1.pdf]

References

1. National Research Corporation's Ticker Survey. 2011. 1 in 5 Americans use social media for health care information URL: http://hcmg.nationalresearch.com/public/News.aspx?ID=9 [accessed 2016-12-23] [WebCite Cache ID 6myWXXPW8]
2. Merriam-Webster. Social Media URL: https://www.merriam-webster.com/dictionary/social%20media [accessed 2016-12-23] [WebCite Cache ID 6myWLjphM]
3. Duggan M. Pew Research Center. 2015. The demographics of social media users URL: http://www.pewinternet.org/2015/08/19/the-demographics-of-social-media-users/ [accessed 2016-12-23] [WebCite Cache ID 6myW3xfFS]
4. Health Research Institute. PricewaterhouseCoopers. 2012. Social media "likes" healthcare: From marketing to social business URL: http://www.pwc.com/us/en/health-industries/health-research-institute/publications/health-care-social-media.html [accessed 2016-12-23] [WebCite Cache ID 6myWDGMXW]
5. Duggan M, Ellison N, Lampe C, Lenhart A, Madden M. Pew Research Center. 2015. Social Media Update 2014 URL: http://www.pewinternet.org/2015/01/09/social-media-update-2014/ [accessed 2016-12-23] [WebCite Cache ID 6myWx1MI9]

6. Facebook. Facebook Help Center. URL: https://www.facebook.com/unsupportedbrowser [accessed 2018-02-18] [WebCite Cache ID 6xJxV7Hzg]
7. YouTube Statistics. URL: https://www.youtube.com/yt/press/statistics.html [accessed 2016-12-23] [WebCite Cache ID 6myX5uS0d]
8. Twitter Company Page. URL: https://about.twitter.com/company [accessed 2016-12-23] [WebCite Cache ID 6myXH4ay9]
9. Grajales FJ, Sheps S, Ho K, Novak-Lauscher H, Eysenbach G. Social media: a review and tutorial of applications in medicine and health care. J Med Internet Res 2014 Feb 11;16(2):e13 [FREE Full text] [doi: 10.2196/jmir.2912] [Medline: 24518354]
10. Terry M. Twittering healthcare: social media and medicine. Telemed J E Health 2009;15(6):507-510. [doi: 10.1089/tmj.2009.9955] [Medline: 19659410]
11. Bosslet GT, Torke AM, Hickman SE, Terry CL, Helft PR. The patient-doctor relationship and online social networks: results of a national survey. J Gen Intern Med 2011 Oct;26(10):1168-1174 [FREE Full text] [doi: 10.1007/s11606-011-1761-2] [Medline: 21706268]
12. Dolan PL. American Medical News. 2011. Nearly all US doctors are now on social media URL: http://www.amednews.com/article/20110926/business/309269969/2/, [WebCite Cache ID 6myXMXrcb]
13. Committee on Professional Liability. Committee Opinion No. 622: Professional use of digital and social media. Obstet Gynecol 2015 Feb;125(2):516-520. [doi: 10.1097/01.AOG.0000460783.32467.bf] [Medline: 25611641]
14. Mostaghimi A, Crotty BH. Professionalism in the digital age. Ann Intern Med 2011 Apr 19;154(8):560-562. [doi: 10.7326/0003-4819-154-8-201104190-00008] [Medline: 21502653]
15. Jain A, Petty EM, Jaber RM, Tackett S, Purkiss J, Fitzgerald J, et al. What is appropriate to post on social media? Ratings from students, faculty members and the public. Med Educ 2014 Feb;48(2):157-169. [doi: 10.1111/medu.12282] [Medline: 24528398]
16. Clyde JW, Domenech RMM, Geiser C. Medical professionalism: an experimental look at physicians' Facebook profiles. Med Educ Online 2014;19:- [FREE Full text] [Medline: 24947922]
17. Ludwig S, Day S. Chapter 7: New standards for resident professionalism: discussion and justification. In: Philibert I, Amis S, editors. The Accreditation Council for Graduate Medical Education (ACGME) 2011 Duty Hour Standards: Enhancing Quality of Care, Supervision, and Resident Professional Development. Chicago, IL: ACGME; 2011:47-51.
18. ABIM Foundation. American Board of Internal Medicine, ACP-ASIM Foundation. American College of Physicians-American Society of Internal Medicine, European Federation of Internal Medicine. Medical professionalism in the new millennium: a physician charter. Ann Intern Med 2002 Feb 05;136(3):243-246. [Medline: 11827500]
19. Johnson KA. The effect of Twitter posts on students' perceptions of instructor credibility. Learning, Media and Technology 2011 Mar;36(1):21-38. [doi: 10.1080/17439884.2010.534798]

Edited by G Eysenbach; submitted 19.05.17; peer-reviewed by L Fernandez-Luque, I Adeleke, K Belcik; comments to author 17.08.17; revised version received 31.12.17; accepted 17.01.18; published 12.03.18.

Please cite as:
Maben-Feaster RE, Stansfield RB, Opipari A, Hammoud MM
Evaluating Patient Perspectives of Provider Professionalism on Twitter in an Academic Obstetrics and Gynecology Clinic: Patient Survey
J Med Internet Res 2018;20(3):e78
URL: http://www.jmir.org/2018/3/e78/
doi:10.2196/jmir.8056
PMID:29530838

Original Paper

Pregnant Women Sharing Pregnancy-Related Information on Facebook: Web-Based Survey Study

Tammy Harpel[1], PhD

Department of Family & Consumer Sciences, Illinois State University, Normal, IL, United States

Corresponding Author:
Tammy Harpel, PhD
Department of Family & Consumer Sciences
Illinois State University
Campus Box 5060
Turner Hall
Normal, IL, 61790-5060
United States
Phone: 1 3094382680
Fax: 1 3094385659
Email: tsharpe@ilstu.edu

Abstract

Background: Research indicates expectant and new mothers use the Internet, specifically social media, to gain information and support during the transition to parenthood. Although parents regularly share information about and photos of their child or children on Facebook, researchers have neither explored the use of Facebook to share pregnancy-related information nor investigated factors that influence such sharing.

Objective: The aim of this study was to address a gap in the literature by exploring the use of Facebook by pregnant women. Specifically, the study examined the use of Facebook to share pregnancy-related information, as well as any association between prenatal attachment and the aforementioned aspects of sharing pregnancy-related information on Facebook.

Methods: Pregnant women who were at least 18 years of age were recruited for participation in the study through posts and paid advertisements on Facebook and posts to professional organization listservs. Individuals interested in participating were directed to a secure Web-based survey system where they completed the consent form and the survey that focused on their current pregnancy. Participants completed the Maternal Antenatal Attachment Scale and answered questions that assessed how often they shared pregnancy-related information on Facebook, who they shared it with, why they shared it, and what they shared.

Results: A total of 117 pregnant women completed the survey. Descriptive statistics indicated that the pregnancy announcement was most commonly shared (75/108, 69.4%), with most women sharing pregnancy-related information on Facebook less than monthly (52/117, 44.4%) with only family and friends (90/116, 77.6% and 91/116, 78.4%, respectively) and for the purpose of involving others or sharing the experience (62/107, 57.9%). Correlation and regression analyses showed that prenatal attachment, in general, was positively and significantly related to all aspects of sharing pregnancy-related information at the $P<.05$ level, with the exception of sharing because of expectations. Quality of attachment, which involves the positive feelings the woman has about her unborn child, was significantly associated with sharing to involve others or share the pregnancy ($t_{8,93}=2.654$, $P=.009$). In contrast, after controlling for other variables, the strength or preoccupation component of prenatal attachment was significantly associated with frequency of sharing ($t_{8,100}=2.554$, $P=.01$), number to types of information shared ($t_{8,97}=2.605$, $P=.01$), number of groups with whom shared ($t_{8,99}=3.467$, $P=.001$), and sharing to get advice ($\chi^2_8=5.339$, $P=.02$).

Conclusions: Pregnant women in this study used Facebook for a variety of reasons, demonstrating the use of the social media platform during pregnancy for supportive and informational purposes. Overall, the results of this study are likely to be useful to professionals who are seeking alternative methods for providing intervention, information, and support to pregnant women via social media in our technology-driven society.

(J Med Internet Res 2018;20(3):e115) doi:10.2196/jmir.7753

KEYWORDS

pregnancy; social media; Facebook

Introduction

Social Media Use and Parents

The transition to parenthood, which begins at pregnancy and continues through the postpartum period, has been altered from decades past by the introduction of new technologies. In particular, medical technologies have provided expectant parents with videos and pictures they can use to introduce their unborn child to others before the birth. Internet technologies have further shaped the transition to parenthood by providing means through which expectant and new parents can share information about their experiences and their unborn child or newborn with family, friends, and acquaintances. According to McDaniel et al, "new mothers appear to be immersed in new age media, such as blogging and social networking," yet, "research on mothers' media use is still in its infancy" [1].

The Pew Research Center began tracking the use of social media sites in 2005 [2]. At that time, 7% of the American population reported using social media. Just 10 years later, 60% of Americans reported they were users of social media sites [2]. Recent statistics indicate there are over 1.5 billion Facebook users [3], with parents comprising a significant portion of social media users in the United States. Of the various social media sites in existence, Facebook is the most common platform used by parents in the United States. In particular, Duggan et al [4] found that 74% of parents reported using Facebook, and Hicks and Brown [5] found that 85% of the pregnant women in their study checked Facebook at least once every day. Of the Facebook-using parents in the Duggan et al [4] study, 75% of parents reported that they logged on to Facebook daily, whereas only 12% reported weekly or less than weekly use of the site. Similarly, Bartholomew et al [6] found that daily Facebook use was common among new mothers. Additional research focusing on the transition to parenthood also shows that new parents use social media on a weekly basis [1].

The Internet, and subsequently social media, has changed the way in which we search for and gather information. In fact, according to Daniels and Wedler, "information seeking through the Internet has become one of the easiest ways to learn about health-related information" [7]. Given the multitude of health-related issues that arise during parenthood, it is not surprising that parents rely on the Internet to gather health-related, as well as parenting information [8]. Pregnant women are no exception when it comes to using the Internet for informational purposes. Research indicates that pregnant women use the Internet to gain reassurance about the normalcy of their pregnancy and symptoms [9] and to gather information about pregnancy, birth, and labor [10,11]. According to Asiodu and colleagues, "social media platforms appeared to be the preferred mechanism for obtaining important information during the antepartum and postpartum periods" for the first-time mothers interviewed in their study [12], with the practice declining during the postpartum period.

Social media also serves a supportive function for expectant and new parents. Duggan et al [4] found that 42% of parents in their study received emotional or social support related to their parenting role through social media. Such support could be gained through online interactions and exchanges with friends and family on social media sites [13] and through membership in social media groups (ie, groups for new mothers). In particular, Tomfohrde and Reinke [8] found that social media fulfilled a supportive function for breastfeeding mothers, whereas new parents in the Thoren et al [14] study reported receiving support through an online support group for premature infants. The tendency of parents to share parenting challenges on Facebook is supported by the finding that social media users were more likely than those who did not use the social media sites to be "aware of stressful events in the lives of their close friends and more distant acquaintances" [15].

Finally, social media appears to facilitate bonding and connections for expectant and new parents [6]. Specifically, Lupton [10] reported that social media provided pregnant women with a sense of connection with their peers, as well as with their own unborn child. The practice of posting status updates and photos on social media appears to play a role in establishing and maintaining connections with others. Researchers have found that new parents shared images of their children on social media [1,10], with over two or three of the new parents in the Bartholomew et al [6] study posting photos of their children each month. In fact, parents in one study reported that posting pictures and comments about their children was their most common social media activity [16]. Prenatal ultrasound technology provides images, "baby's first picture" [17], that can be shared prenatally by expectant parents, and research indicates that pregnant women share the ultrasound images to facilitate bonding and to involve others in their pregnancy [18]. Saetnan identified ultrasound technology as a "family-building technology," noting its ability to involve others in the pregnancy, foster supportive interactions, and initiate "thinking about the baby as a family member" before birth [19]. Facebook and other social media platforms provide additional opportunities for expectant parents to share the technologically produced ultrasound images for these purposes, thus augmenting the notion of "family-building technology" to include social media. In fact, Johnson suggested that "Facebook may be one of the social communities in which women and their partners first announce their pregnancy and where they share ultrasound images, their experiences during pregnancy as well as their excitement at the impending arrival of their baby" [20].

Parental use of social media sites appears to vary by pregnancy status, gender, and age of the parent. In particular, women pregnant for the first time are more likely to use social media than multiparous women [21]. As is the case with social media use in general [2], younger parents (younger than 40 years) are more likely to use social media [22] and Facebook [4] than older parents. There are also gender differences in terms of which social media site is being used by parents, with mothers using Facebook significantly more than fathers for support (80% mothers, 65% fathers) and informational purposes (83% mothers, 74% fathers) [4]. Recently, Bartholomew et al [6] explored Facebook use during the postpartum period in relation to the parenting role, finding that new mothers were more likely than new fathers to utilize the social networking site.

Prenatal Attachment

As the earliest conceptualization, the definition of prenatal attachment has evolved from encompassing maternal behaviors indicative of the mother's affiliation toward and interaction with her unborn child [23], to later becoming a multidimensional construct that also involves expectant parents' thoughts and fantasies about the fetus [24]. Doan and Zimmerman [25,26] further conceptualized prenatal attachment as involving multiple components—behavioral, cognitive, and affective. More recently, Krisjanous et al defined prenatal attachment as "the emotional attachment made up of feelings of affiliation and affection by the mother to the developing baby, which indicates positive acceptance and acknowledgment that the pregnancy is producing a person in their own right" [27].

Most relevant to this study, Condon and Corkindale [28] conceptualized prenatal maternal attachment as "a desire for knowledge about the fetus, pleasure in interaction with the fetus (both in fantasy and reality), and a desire to protect the unborn baby and his/her own needs, even at the expense of the mother's own" [28]. More specifically, they conceptualized it as comprising two components—the quality of the pregnant woman's attachment to and her preoccupation with her unborn child. The quality of prenatal attachment involves the pregnant woman's positive feelings for and clear images of her unborn child, whereas the preoccupation component involves the strength of the pregnant woman's attachment to her unborn child. According to Condon and Corkindale [28], the strength of her attachment is manifested through such processes as time spent thinking and talking about the unborn child and attempts to engage in behaviors that are healthy and protective of her unborn child.

Despite the various conceptualizations of the construct, research consistently indicates that prenatal attachment increases over the course of the pregnancy [29,30], with moderate stability carrying into toddlerhood [31]. Multiple factors have been investigated in relation to prenatal attachment, with research indicating that prenatal attachment is higher among women who are pregnant for the first time [32] and those who are involved in a supportive couple relationship [33,34]. In addition, certainty about fetal sex [29], quickening, and prenatal technology [35-38] have also been associated with maternal prenatal attachment.

Although parent-child attachment, during the postpartum period and later, typically receives more attention from researchers and interventionists, research has highlighted the importance of attachment during the prenatal period. In particular, researchers [33] found a positive association between maternal-fetal attachment and engagement in self-care behaviors and positive health practices during pregnancy. Furthermore, prenatal attachment also appears to be associated with cigarette smoking, a behavior that poses serious risks to the pregnancy and newborn [39]. Specifically, researchers found a negative correlation between maternal prenatal attachment and the number of cigarettes smoked by the woman during the pregnancy [40], as well as higher presence of particular elements of prenatal attachment among women who quit smoking during pregnancy versus those who did not [41]. Similarly, Ross [42] found a negative correlation between prenatal attachment and alcohol consumption during pregnancy. Overall, the research indicates that prenatal attachment is negatively associated with harmful maternal behaviors and positively associated with behaviors that are more optimal during pregnancy. It is then not surprising that prenatal attachment has also been associated with negative neonatal outcomes such as low birth weight [33,43].

This Study

Although the associations between prenatal attachment and maternal behaviors during pregnancy have been investigated, the role of prenatal attachment in the mother's representation of her unborn child to others has not been investigated. Furthermore, despite the importance of Facebook in parents' lives, research on expectant parents' use of Facebook to share pregnancy-related information is lacking. More specifically, to date, researchers have not investigated a potential relationship between prenatal maternal attachment and pregnancy-related posts on social media. On the basis of the findings of previous research and the components of attachment, as identified by Condon and Corkindale [28], it seems plausible that a relationship does exist between prenatal attachment and social media posts about one's pregnancy and unborn child. In terms of the strength (preoccupation) component of prenatal attachment, one may assume that pregnant women who are more attached to their unborn child would post about their pregnancy and unborn child more frequently to Facebook than those who are less attached to their unborn child, as they are spending more time thinking about their unborn child. In addition, the avoidance of harmful behaviors, an additional aspect of the strength component of prenatal attachment, may embolden a pregnant woman to seek advice and support regarding her pregnancy and unborn child's health from others on social media. Finally, a pregnant woman who has clearer images of and more positive feelings about her unborn child, both indicators of the quality of prenatal attachment, would likely be more compelled to post about her pregnancy and unborn child in attempts to involve others in her pregnancy (to share in her excitement) and to introduce her unborn child as a member of the family. Taken together, therefore, it is hypothesized that pregnant women utilize social media, specifically Facebook, to share pregnancy-related information with others and that such sharing is related to the pregnant woman's prenatal attachment toward her unborn child.

Specifically, the study examined the use of Facebook to share information about pregnancy among pregnant women, as well as any association between prenatal attachment and sharing pregnancy-related information on Facebook.

The following research questions (RQs) were explored in the study:

- RQ1. What pregnancy-related information are pregnant women sharing on Facebook?
- RQ2. With whom are they sharing pregnancy-related information on Facebook?
- RQ3. How often do they share pregnancy-related information on Facebook?
- RQ4. Why do they share pregnancy-related information on Facebook?

In addition, the following hypotheses (Hs) were tested in the study:

- H1. Prenatal attachment will be associated with frequency of sharing pregnancy-related information on Facebook.
- H2. Prenatal attachment will be associated with how many types of pregnancy-related information are shared on Facebook
- H3. Prenatal attachment will be associated with the number of groups with whom pregnancy-related information is shared on Facebook
- H4. Prenatal attachment will be associated with the reasons for sharing pregnancy-related information on Facebook.

Methods

Recruitment

Upon receiving Human Subject Approval, pregnant women who were at least 18 years of age were recruited for participation in the study through posts to the researcher's personal and research Facebook pages that were subsequently shared by others, paid advertisements on Facebook that were targeted to pregnancy-related groups and pages, and posts to listservs of the researcher's national professional organization. Individuals interested in participating were directed through the announcement to the secure Web-based survey system utilized by the university where they completed the consent form and the survey. In addition to the consent form, the survey consisted of one page of demographic questions, one page of questions assessing the use of Facebook to share pregnancy-related information, and one page of questions comprising the attachment scale for a total of 41 questions. To facilitate completion and expedite movement through the survey, question condition settings were included to automatically skip questions that were not relevant to participants. Due to university human subject stipulations, participants were allowed to exit the survey at any time and skip any questions in the survey with the exception of indicating their consent, or lack thereof, to participate on the consent form. Participants were instructed to complete the survey with their current pregnancy as the focus of their answers. The survey was open and set for single response submission, without the capability to update responses after submission; although, participants were able to return to previous questions before submission.

Before data collection, the Web-based survey was pilot tested by four women known to the researcher—two of whom were pregnant and two who had recently given birth. Issues with survey formatting and word choice were resolved based on feedback provided from the pilot participants. In addition, the pilot participants provided the researcher with information concerning the time required for survey completion, which was subsequently used for the time-to-complete estimate provided on the consent form.

Measures

Dependent Variables

Frequency of sharing information, types of information shared, with whom information was shared, and reasons for sharing information were the dependent variables in this study. Each was measured with closed-ended questions on the survey.

Frequency of Sharing Information

Participants indicated how often they posted pregnancy-related information to Facebook during the current pregnancy by choosing one of the following options: (1) less than once per month, (2) once per month, (3) a few times per month, (4) once per week, (5) a few times per week, (6) once per day, or (7) more than once per day. The options were developed by the researcher using categories from the Bartholomew et al [6] study as a framework. A higher number represented more frequent sharing of pregnancy-related information on Facebook.

Types of Information Shared

Participants indicated which of the following types of pregnancy-related information they had shared on Facebook during their current pregnancy by indicating "yes" or "no" for each type: (1) announcement of their pregnancy, (2) ultrasound pictures or videos of their unborn child, (3) announcement of their unborn child's sex, (4) information about their pregnancy symptoms, (5) information about their preparation for the baby, (6) information about medical appointments, (7) information about the progression of the pregnancy, (8) information about pregnancy complications, and (9) information about the birth plans. These categories were developed by the researcher based on the researcher's previous research with expectant parents and personal correspondence with pregnant women about the types of pregnancy-related information they shared with others. Each item was coded, with a 1 indicating the participant shared the information and 0 indicating the participant did not share the information on Facebook. Scores on the nine items were summed, with a higher score indicating the sharing of more types of pregnancy-related information on Facebook.

Whom Shared With

Participants indicated which of the following groups they shared pregnancy-related information with on Facebook: (1) only family, (2) only friends, (3) only friends and family, or (4) public (no restrictions on who could see the information). These categories were developed by the researcher, with the privacy settings available to Facebook users (public, friends, and customize) serving as the initial framework for the categories. A "yes" response to each category was coded as 1 and a "no" response coded as 0. Scores for the four groups were summed to gain a *whom shared with* score (ranging from 0-4), with a higher number indicating more groups with whom the information was shared.

Reasons for Sharing

Participants indicated the reasons they posted pregnancy-related information to Facebook by indicating which of the following options applied to them: (1) to share excitement, (2) to document pregnancy, (3) to get advice, (4) to involve others in the pregnancy, (5) to issue a prayer request, and (6) others expected them to share it. The response options were developed by the researcher, utilizing reasons cited in literature and from personal correspondence with pregnant women. A response of "yes" for an item was coded as 1, whereas a "no" response was coded as 0. Subsequent factor analysis was performed on the six reasons

for sharing information, with the following four categories of reasons identified: (1) getting advice, (2) issue prayer request, (3) involve others or share experience, and (4) expected to.

Independent Variables

Prenatal attachment served as the independent variable in the study. In addition, demographic variables were treated as controls in the analyses.

Prenatal Attachment

Prenatal attachment was assessed with the 19-item Maternal Antenatal Attachment Scale (MAAS) [28,44]. The MAAS consists of two subscales. The first subscale, which is comprised of 10 items, assesses the pregnant woman's *quality* of attachment to the unborn child (ie, her positive feelings about the unborn child and clear mental images of the unborn child), whereas the second 8 item *preoccupation* subscale assesses the strength of the pregnant woman's attachment with her unborn baby (eg, the amount of time she thought about the unborn baby and protective behaviors). Examples of questions from the quality subscale are "Over the past two weeks when I think about the baby inside me I get feelings which are (very sad, moderately sad, a mixture of happiness and sadness, moderately happy, very happy)" and "The picture in my mind of what the baby at this stage actually looks like inside the womb is (very clear, fairly clear, fairly vague, very vague, I have no idea at all)." Items from the preoccupation subscale include "Over the past two weeks I have had dreams about the pregnancy or baby (not at all, occasionally, frequently, very frequently, almost every night)" and "Over the past two weeks I have taken care with what I eat to make sure the baby gets a good diet (not at all, once or twice when I ate, occasionally when I ate, quite often when I ate, every time I ate)." Per the instrument guidelines, one question ("Over the past two weeks I have felt that the baby inside me is dependent on me for its well-being") was included in the overall attachment score, but not in either of the subscales. Participants indicated their level of agreement with each statement on a 5-point Likert scale, with a total computed for the full scale and each subscale, and higher totals indicating greater prenatal attachment.

Control Variables

Participants provided the following information: (1) age (5 categories), (2) parity status (first pregnancy vs not first pregnancy), (3) knowledge of fetal sex (yes or no), (4) weeks currently pregnant (6 categories), and (5) planned pregnancy (yes, no). Specific categories for the control variables that appeared on the survey are listed in Table 1.

Statistical Analysis

Statistical Package for the Social Sciences (SPSS) version 23 (IBM Corp) was used for all analyses. Descriptive statistics were used to analyze the data for research questions 1 to 4. Correlations (Pearson, Spearman rho) were used to test hypotheses 1 to 4, with multiple and binomial logistic regression used to test for associations between the independent and dependent variables when controlling for the other variables and demographic variables (participant's age, first pregnancy status, knowledge of fetal sex, number of weeks pregnant, and planning of the pregnancy). Only the attachment subscales (preoccupation and quality), and not the overall prenatal attachment scores, were included in the regression analyses to reduce multicollinearity between the overall scale and subscale scores.

Results

Participant Characteristics

There were 5395 clicks on the survey link during the 7-month data collection period. From those clicks, 218 individuals consented to participate in the study, and a total of 117 pregnant women completed the Web-based survey. The sample was predominately white (90/109, 82.6%), married (85/110, 77.3%), college educated (74/113, 65.5%), and in the age range of 26 to 29 years (42/114, 36.8%). In addition, almost half of the women were pregnant for the first time (56/113, 49.6%), most of the pregnancies were planned (76/113, 67.3%), and the most common category for weeks pregnant was 27 to 33 weeks (46/113, 40.7%). See Table 1.

The mean on the overall attachment scale was 75.68 (SD 7.655; range: 49-88.70). The mean on the preoccupation (strength) subscale was 28.66 (SD 4.967; range: 15-38.70), and the mean for the quality subscale was 42.38 (SD 3.569; range: 20.41-47).

Research Questions

Research Question 1

The most common type of information shared on Facebook was the pregnancy announcement (75/108, 69.4%), followed by sharing pregnancy progress (57/110, 51.8%), and announcing the fetus' sex (55/111, 49.5%). The least common type of information shared was birth plans (7/116, 6.0%). See Table 2 for more complete results.

Research Question 2

Friends and family were the most common recipients of the pregnancy-related information posted to Facebook. Over 75% of the women shared information with friends (91/116, 78.4%) and family (90/116, 77.6%). The women were less likely to share with individuals they did not know personally (Table 2).

Research Question 3

The majority of the participants posted pregnancy-related information relatively infrequently, with 44.4% (52/117) indicating they posted information related to their current pregnancy less than once per month, followed by sharing a few times per month (18.8%, 22/117). A smaller percentage of women shared information a few times per week, with even fewer sharing pregnancy-related information on a daily basis (Table 2).

Research Question 4

When considering all six of the reasons for sharing, the most common reason was to share the excitement of the pregnancy with others (57.9%, 62/107), followed by the desire to document the pregnancy (31.3%, 35/112), and get advice (28.9%, 33/114). The least common reason for sharing was feeling pressured by others to share the information (Table 2), indicating that most of the women voluntarily shared the information with others.

Table 1. Demographic information.

Demographic characteristics	n (%)
Age, years	
18-21	12 (10.7)
22-25	19 (16.7)
26-29	42 (36.8)
30-33	29 (25.4)
34-39	9 (7.9)
40-44	1 (0.9)
≥45	0 (0.0)
Race	
White	90 (82.6)
Black	8 (7.3)
Hispanic or Latino	4 (3.7)
Native American	1 (0.9)
Asian or Pacific Islander	5 (4.6)
Other	1 (0.9)
Annual income (USD)	
<30,000	25 (22.7)
30,000-49,999	17 (15.5)
50,000-74,999	20 (18.2)
75,000-99,999	18 (16.4)
>100,000	30 (27.2)
Education	
Less than high school	3 (2.7)
High school or general equivalency diploma	14 (12.4)
Some college or vocational training	22 (19.5)
Associate degree	3 (2.7)
Bachelor’s degree	37 (32.7)
Master’s degree	23 (20.4)
Doctorate degree	11 (9.7)
Marital status	
Single, never married	19 (17.3)
Divorced	3 (2.7)
Separated	2 (1.8)
Widowed	1 (0.9)
Married	85 (77.3)
Weeks pregnant	
Under 13 weeks	5 (4.4)
13 to 19 weeks	16 (14.2)
20 to 26 weeks	16 (14.2)
27 to 33 weeks	46 (40.7)
34 to 40 weeks	29 (25.7)
Over 40 weeks	1 (0.9)

Table 2. Sharing pregnancy-related information.

Category of sharing behavior	n (%)
Type of information shared	
Pregnancy announcement	75 (69.4)
Pregnancy progress	57 (51.8)
Sex of fetus	55 (49.5)
Ultrasound pictures	46 (39.3)
Pregnancy symptoms	36 (32.1)
Preparation for baby	33 (29.2)
Medical appointments	17 (15.0)
Pregnancy complications	14 (12.2)
Birth plans	7 (6.0)
With whom shared	
Friends	91 (78.4)
Family	90 (77.6)
Friends of friends	15 (12.9)
Everyone	7 (6.0)
Frequency shared	
Less than monthly	52 (44.4)
Once per month	11 (9.4)
Few times per month	22 (18.8)
Once per week	8 (6.8)
Few times per week	16 (13.7)
Once per day	6 (5.1)
More than once each day	2 (1.7)
Reasons for sharing	
Share excitement	62 (57.9)
Document pregnancy	35 (31.3)
Get advice	33 (28.9)
Involve others	19 (17.0)
Issue prayer request	18 (15.9)
Others expected it	9 (7.8)

Hypotheses

Hypothesis 1

Results of Pearson's correlations indicated that overall prenatal attachment and the quality and preoccupation subscales were positively and significantly associated with *frequency of sharing* (Table 3). Participants who reported higher overall prenatal attachment were more preoccupied with their unborn child and had more positive feelings about their unborn child shared information about their current pregnancy more frequently on Facebook.

Using the enter method of multiple regression, the control variables and quality and preoccupation subscale scores were entered in the equation. Results indicated that the age of participants was found to be a significant predictor, and the preoccupation subscale retained significance after controlling for other variables (Table 4). However, the quality subscale of attachment was no longer a significant predictor of frequency of sharing information. Participants who were younger and more preoccupied with their unborn baby shared pregnancy-related information more frequently on Facebook. Therefore, H1 was partially supported as the preoccupation subscale of attachment was found to be significantly associated with frequency of sharing pregnancy-related information on Facebook after controlling for other variables.

Hypothesis 2

Overall, prenatal attachment and the two subscales (preoccupation and quality) were significantly associated with

sharing *more types* of pregnancy-related information on Facebook (Table 3). Specifically, participants who reported higher overall prenatal attachment, higher quality attachment, and more preoccupation with their unborn child shared more types of pregnancy-related information on Facebook.

The enter method of multiple regression was used to test the associations after controlling for other variables. Results indicated that the preoccupation subscale, age, and weeks pregnant were significant contributors to the model, with those who were more preoccupied with their unborn child, younger, and further along in their pregnancies sharing more types of information (Table 5). The quality of attachment was no longer significant after controlling for the influence of other variables. Therefore, H2 was partially supported with the preoccupation subscale of attachment being significantly and positively related to the total types of pregnancy-related information shared on Facebook.

Hypothesis 3

Analysis with Pearson correlation revealed that overall prenatal attachment and the preoccupation subscale were significantly associated with sharing pregnancy-related information with *more groups* on Facebook (Table 3). Specifically, participants who reported higher overall prenatal attachment and were more preoccupied with their unborn baby shared pregnancy-related information with more groups of people on Facebook.

The enter method of multiple regression model was again used to test associations between prenatal attachment and number of groups with whom information was shared while controlling for other variables. Age and preoccupation with the unborn baby were significant contributors to the model (Table 6). H3 was partially supported in that participants who were more preoccupied with their unborn baby (a subscale of attachment) shared pregnancy-related information with significantly more groups of people on Facebook.

Table 3. Correlations between independent and dependent variables.

Variables	Prenatal attachment	Preoccupation attachment subscale	Quality attachment subscale
Frequency shared	.338[a]	.308[a]	.228[a]
Total types shared	.332[a]	.360[a]	.194[b]
Groups shared with	.248[a]	.338[a]	.050
Share pregnancy	.338[a]	.274[a]	.303[a]
Get advice	.236[b]	.277[a]	.091
Prayer request	.251[a]	.238[a]	.163
Expected to	−.135	−.096	−.177

[a]Correlation significant at the .01 level.

[b]Correlation significant at the .05 level.

Table 4. Multiple regression results for hypothesis 1 (frequency of sharing; N=108; R^2=.227, $F_{8,100}$=3.671, and P<.001).

Variable	B (SE)	Beta	*t* value[a]	*P* value
Constant	1.349 (2.366)	N/A	0.570	.57
Controls				
Age	−.488 (0.149)	−.336	−3.286	.001
Number of children	.358 (0.230)	.194	1.556	.12
Fetal sex	−.428 (0.433)	−.105	−0.98	.32
Weeks pregnant	−.050 (0.079)	−.065	−0.625	.53
Planned pregnancy	.012 (0.350)	.003	0.035	.97
First pregnancy	.438 (0.416)	.131	1.055	.29
Predictors				
Preoccupation (attachment)	.090 (0.035)	.262	2.554	.01
Quality (attachment)	−.010 (0.054)	−.021	−0.185	.85

[a]Degrees of freedom for *t* test values=8,100.

Table 5. Multiple regression results hypothesis 2 (number of types of information shared; N=105; R^2=.314, $F_{8,97}$=5.540, P<.001).

Variable	B (SE)	Beta	*t* value[a]	*P* value
Constant	−2.815 (3.308)	N/A	−.851	.4
Controls				
Age	−.717 (0.208)	−.337	−3.454	.001
Number of children	.474 (0.322)	.174	1.471	.14
Fetal sex	−.735 (0.610)	−.121	−1.206	.23
Weeks pregnant	.328 (0.111)	.291	2.940	.004
Planned pregnancy	.378 (0.500)	.072	0.756	.45
First pregnancy	.552 (0.581)	.112	0.949	.35
Predictors				
Preoccupation (attachment)	.129 (0.050)	.254	2.605	.01
Quality (attachment)	.038 (0.076)	.053	0.498	.62

[a]Degrees of freedom for *t* test values=8,97.

Table 6. Multiple regression hypothesis 3 (groups with whom information shared; N=107; R^2=.252, $F_{8,99}$=4.159, P<.001).

Variable	B (SE)	Beta	*t* value[a]	*P* value
Constant	1.570 (2.344)	N/A	0.670	.51
Controls				
Age	−.348 (0.150)	−.233	−2.313	.02
Number of children	.137 (0.234)	.072	0.585	.56
Fetal sex	−.127 (0.442)	−.030	−0.287	.77
Weeks pregnant	−.052 (0.080)	−.066	−0.641	.52
Planned pregnancy	−.586 (0.355)	−.161	−1.649	.10
First pregnancy	.175 (0.421)	-.051	-.415	.68
Predictors				
Preoccupation (attachment)	.125 (0.036)	.350	3.467	.001
Quality (attachment)	−.016 (0.055)	−.032	−0.285	.77

[a]Degrees of freedom for *t* test values=8,99.

Hypothesis 4

Pearson correlation results indicated that overall prenatal attachment and the preoccupation and quality of attachment subscales were significantly and positively associated with sharing information to *involve others or share the pregnancy experience* (Table 3) . Results of the multiple regression analysis indicated that after controlling for the other variables, age, weeks pregnant, and quality of attachment were significantly associated with sharing to involve others (Table 7). The preoccupation subscale was no longer significant after controlling for the other variables. Participants who were younger, further along in their pregnancy, and reported higher quality of attachment were significantly more likely to share pregnancy-related information to involve others or share the pregnancy experience with others.

Spearman rho correlational analysis indicated that, overall, prenatal attachment and the preoccupation attachment subscale were significantly and positively associated with sharing to *get advice* from others (Table 3). Binomial logistic regression analysis indicated that after controlling for other variables, the preoccupation attachment subscale retained significance (Table 8), meaning that individuals who reported greater preoccupation with their unborn baby were significantly more likely to post pregnancy-related information to Facebook in order to get advice from others.

Results of Spearman rho analysis showed that overall prenatal attachment and the preoccupation attachment subscale were significantly, positively associated with sharing pregnancy-related information to *issue a prayer request* (Table 3). However, after controlling for other variables using binomial logistic regression, the model was not significant, χ^2_8=3.5, N=108, P=.90, nor were the independent or control variables. Thus, prenatal attachment was not significantly associated with sharing to issue a prayer request after controlling for other variables.

Table 7. Multiple regression results hypothesis 4 (sharing to involve or share experience; N=101; R^2=.229, $F_{8,93}$=3.460, *P*=.002).

Variable	B (SE)	Beta	*t* value[a]	*P* value
Constant	−4.427 (1.998)	N/A	−2.216	.03
Controls				
Age	−.319 (0.126)	−.266	−2.533	.01
Number of children	.172 (0.192)	.114	0.896	.37
Fetal sex	−.037 (0.372)	−.011	−0.099	.92
Weeks pregnant	.146 (0.070)	.229	2.084	.04
Planned pregnancy	−.097 (0.297)	−.327	−0.327	.74
First pregnancy	.049 (0.350)	.139	0.139	.89
Predictors				
Preoccupation (attachment)	.034 (0.030)	.119	1.132	.26
Quality (attachment)	.121 (0.046)	.304	2.654	.009

[a]Degrees of freedom for *t* test values=8,93.

Table 8. Logistic regression results for hypothesis 4 (sharing to get advice; N=108; χ^2_8=7.5, *P*=.48).

Variable	B (SE)	Wald chi-square[a]	*P* value
Constant	−2.439 (3.595)	0.5	.497
Controls			
Age	−.325 (0.237)	1.9	.17
Number of children	−.135 (0.376)	0.1	.72
Fetal sex	−.404 (0.666)	0.4	.54
Weeks pregnant	.197 (0.130)	2.3	.13
Planned pregnancy	.504 (0.539)	0.9	.35
First pregnancy	.359 (0.654)	0.3	.497
Predictors			
Preoccupation (attachment)	.138 (0.060)	5.3	.02
Quality (attachment)	−.056 (0.087)	0.4	.52

[a]Degrees of freedom for Wald chi-square=8.

Finally, results of Spearman rho analysis indicated that none of the attachment scores were significantly associated with sharing because of the *expectations of others*. Therefore, the overall prenatal attachment score and the quality and preoccupation subscales were not significantly associated with sharing pregnancy-related information on Facebook to conform to others' expectations to do so.

Discussion

Principal Findings and Comparison With Prior Work

The women who participated in the study reported sharing pregnancy-related information relatively infrequently on Facebook. In fact, the majority of the women reported sharing the information less than monthly, with less than 7% of the women sharing information on a daily basis. This finding contradicts that of previous research that indicated 75% of parents [4] and 85% of pregnant women [5] checked Facebook on a daily basis. However, this discrepancy in results may be because of the fact that this study did not explore the frequency of general Facebook use (ie, using Facebook for purposes other than sharing pregnancy-related information). Instead, the participants were asked only how frequently they shared pregnancy-related information. Therefore, it may be the case that the participants did log on to Facebook more frequently for general use than they did to post pregnancy-related information. Furthermore, it is likely that pregnancy-related information is shared on Facebook relatively infrequently because the salience of the information is infrequent. When there is something novel to post, pregnant women post it; however, such novel information is likely to be a one-time occurrence or relatively infrequent. In support of this explanation, two of the three most commonly shared types of information by women in this study reflected one-time occurrences (announcing the pregnancy and sharing the fetus' sex).

The findings of this research also showed that the women voluntarily shared pregnancy-related information with a rather select group of individuals, namely their friends and family on Facebook. Overall, very few shared because of the expectations of others, and few shared with individuals they did not know (ie, friends of friends and everyone on Facebook). The most common reasons for sharing the information fit the overall category of sharing to involve others or share the excitement of the pregnancy. By sharing information about their pregnancy, unborn child, and preparations, they were providing others access to their pregnancy experience and enhancing connections with others during their pregnancy in their role as an expectant mother. Therefore, the findings appear to support researcher assertions that social media use facilitates connection for pregnant women [6,10], and Facebook specifically may provide new parents opportunities to maintain social ties during the transition to parenthood [6].

Aside from the aforementioned reasons, posting to gain advice from others was the third most common reason for sharing pregnancy-related information on Facebook, with one-third of the sample reporting this reason. This finding corresponds to previous research documenting the value of social media as an information-gathering tool [7] and the practice of pregnant women seeking information and advice through social media [9-11]. Unfortunately, the actual content of the women's posts to Facebook was not investigated in this study, thus leaving the specific types of advice they were seeking unknown. However, given that pregnancy symptoms were shared by over one-third of the sample, it is possible that a portion of the advice they sought was related to the symptomology of pregnancy, as well as advice related to the other types of information shared by the women (ie, preparing for baby, complications, and birth plans).

The four types of pregnancy-related information most commonly shared on Facebook by the women in this study were the pregnancy announcement, information about the progression of the pregnancy, the sex of their unborn child, and ultrasound photos. In general, these findings support Johnson's [20] assertion that Facebook provides expectant parents with opportunities to announce their pregnancy and share images of their unborn baby with others. In terms of specific findings, 69% of the pregnant women in this study reported they had announced their pregnancy via Facebook, whereas 39% reported they had shared an ultrasound image of their unborn child on Facebook. The latter finding coincides with previous research in which sharing comments and photos of one's child or children was the most common activity carried out by parents on Facebook [16]. Although the percentage of pregnant women who shared an ultrasound image in this study is lower than that reported among new parents sharing photos in the Bartholomew et al [6] study, the difference may be because of a lack of access to a quality ultrasound image, whereas new parents typically accrue many photos of their newborn that are suitable and available for sharing. Given that the study participants did not report on their possession of ultrasound images, this explanation is merely speculative. However, the Bartholomew et al [6] finding that new parents reported uploading and posting more pictures of their child during the postpartum period than they did during the pregnancy lends support to this explanation.

In general, the findings regarding sharing pregnancy-related information on Facebook provide support for the family-building capabilities of technology. Although past research indicated that prenatal technology assisted with introducing an unborn child as a member of the family and building an identity for the unborn child in the family [17,19], our findings indicate that social media also serves as a family-building tool during the prenatal period. The pregnant women in this study voluntarily shared information about their pregnancy and preparations with others to facilitate involvement in the pregnancy experience and, most likely, to establish connections. In addition, the women shared specific information about their unborn child on Facebook, specifically the unborn child's sex and ultrasound image. Thus, by way of sharing information, their unborn child could establish an identity before birth, as previously suggested by Johnson [20], among those who viewed the posts. In particular, the image could allow others to explore the child's physical features for family resemblance, while information about the unborn child's name, prenatal habits (ie, kicks), and sex could facilitate the development of the child's role within the family [19] and speculations about his or her future personality, interests, and behaviors.

This is the first study to investigate the role of prenatal attachment with respect to pregnant women sharing information about their pregnancy on Facebook, with results indicating that prenatal attachment is positively related to sharing pregnancy-related information. To begin, results of correlational analyses showed that prenatal attachment, in general, was positively and significantly related to all aspects of sharing pregnancy-related information, with the exception of sharing because of expectations. Thus, pregnant women who were more attached to their unborn child were more likely to voluntarily post information related to their pregnancy or unborn child on Facebook. More specifically, before controlling for other variables, the preoccupation component of prenatal attachment was significantly and positively associated with all aspects of sharing, again with the exception of sharing because of others' expectations. This finding makes intuitive sense. One would expect that a pregnant woman who possesses a stronger attachment to her unborn child, meaning she spends more time thinking about and has stronger feelings for her unborn child, would post more pregnancy-related information on Facebook than a pregnant woman who spends less time preoccupied with her unborn child.

These findings may also be interpreted in relation to the concept of maternal identity or, more specifically, maternal identify confirmation. According to Allen and Hawkins, maternal identity confirmation is the "desire for the external validation of the maternal role" [45]. Recently, this concept was investigated in relation to Facebook use among parents. Specifically, Schoppe et al [46] found positive relationships between maternal identity confirmation and aspects of Facebook activity. In particular, women who sought more confirmation of their maternal identity were more likely to post photos of their child and to use their child's photo as their own profile picture on Facebook. Drawing on Schoppe et al's findings, it may be that salience of one's maternal identity played a role in the sharing of pregnancy-related information on Facebook in

this study. More specifically, pregnant women for whom the maternal identify was more salient may have posted more pregnancy-related information on Facebook to receive validation of their maternal role and affirmation "that they are doing motherhood (in this case, pregnancy) correctly and normatively" [46].

Overall, the findings of this study indicate that, when considering the two subscales of attachment used in this study, the *strength*, or preoccupation component, of prenatal attachment is associated with more aspects of sharing pregnancy-related information on Facebook than is the *quality* component of prenatal attachment. Before controlling for other variables in the regression analyses, the quality subscale was associated with three aspects of sharing, whereas the preoccupation subscale was associated with six aspects of sharing. More specifically, after controlling for other variables, quality of attachment, which involves the positive feelings the woman has about her unborn child, was significantly associated with only sharing to involve others or share the pregnancy. In contrast, the strength or preoccupation component of prenatal attachment was significantly associated with frequency of sharing, number of types of information shared, number of groups with whom information was shared, and sharing to get advice. The question arises, "Why is strength of prenatal attachment significantly related to these aspects of sharing when quality of the attachment is not?" Obviously, a causal influence cannot be determined by the methodology used in the study. However, the explanation may be as simple as women who are more preoccupied with their unborn child or pregnancy manifest that preoccupation by sharing more types of information about their pregnancy more frequently and with more people. In addition, their preoccupation may involve anxiety about the pregnancy or unborn child that, in turn, prompts them to seek advice, validation, or reassurance on Facebook. Research on adult attachment and Facebook activity may provide support for this explanation in that adults who are more anxious about their relationships (ie, anxiously attached) are more likely to post about and seek visibility of their romantic relationships on Facebook [47]. However, this evidence should be viewed with caution, given that adult attachment and prenatal attachment are different constructs, as the former focuses on the adult within romantic and intimate relationships [48], whereas the latter focuses on the adult's thoughts, feelings, and behaviors toward or about their unborn child and hence not within the context of a bidirectional relationship. That said, although one may argue that the applicability of adult attachment research to this study is questionable, it does shed some light on the potential role of anxiety in posting of pregnancy-related information on Facebook, particularly in relation to the strength or preoccupation component of prenatal attachment.

Similarly, it may also be that the pregnant women who were more preoccupied with their unborn child possess a personality trait, or other characteristic, that played a role in their sharing behavior, as well as their tendency to be more preoccupied with their unborn child. Neuroticism, which involves anxiety and worry, is one such trait that may moderate the relationship between prenatal attachment and sharing of pregnancy-related information on Facebook. Although research on personality types and social media use during pregnancy is lacking, there is evidence to suggest that the personality trait of neuroticism is positively related to Facebook activity among adults [49,50]. With respect to parenting, although Schoppe-Sullivan et al [46] did not find significant relationships between personality types and frequency of Facebook activity among parents, they did find that parents who scored higher on neuroticism posted pictures of their child sooner after the child's birth. Therefore, it could be that pregnant women who are also neurotic are more preoccupied about their unborn child and, in turn, use Facebook as a platform to solicit information, reassurance, and support. Future research is needed to refute or support this potential moderating effect.

Though not a specific focus of this research, findings related to the control variables are worthy of mention. Specifically, in the regression analyses, weeks pregnant and age were significantly associated with aspects of sharing pregnancy-related information on Facebook. Age was negatively and significantly associated with frequency of sharing, number of types of information shared, number of groups with whom information was shared, and sharing to involve others. These findings confirm prior research that indicated a negative correlation between Facebook use and age among the general population and parents [4,2]. Furthermore, weeks pregnant, another control variable in this study, was found to have a significant positive association with aspects of sharing pregnancy-related information on Facebook. Participants who were further along in their pregnancy shared more types of information and were more likely to share to involve others in the pregnancy. This finding is not surprising given that, as the pregnancy progresses, there are more opportunities and more types of pregnancy-related information to share (ie, more frequent appointments, preparations for birth, etc), and prenatal attachment has been shown to increase throughout gestation [29,30]. Another worthy explanation may involve Rothman's [51] notion of the "tentative pregnancy" in which women suspended attachment and excitement about their pregnancy until they were assured of the viability and health of the fetus through prenatal testing. In particular, and of relevance to this study, the women interviewed by Rothman postponed announcing their pregnancy until they received the results of prenatal testing that confirmed the fetus' health or their continuation of their pregnancy. More recently, Ross [52] asserted that the concept of "tentative pregnancy" could be extended to pregnancy in general, not only to instances of genetic testing. Thus, the women in this study may have delayed or suspended posting about their pregnancy until they were more assured of the health and viability of their pregnancy and unborn child. Given that our society perceives the first trimester as a period of risk and many discourage announcing a pregnancy during this period, it is worth noting the possibility that the "tentative pregnancy" may have played a role in the women's posting behavior and, subsequently, the findings of this study. The women did not report *when* they posted about the pregnancy on Facebook; therefore, it is not possible to ascertain if this explanation is accurate or fitting of the women in this study. However, women in Ross' research "engaged with the convention of keeping news of their pregnancy secret during its early stages" [52]. Therefore, it seems plausible that the women in this study succumbed to this practice as well.

Limitations and Suggestions for Future Research

Despite its contributions to existing literature, this study was not without limitations. To begin, the sample size was relatively small and homogenous with regard to demographic characteristics. This may have impacted the strength of the relationships between the variables, while also limiting the generalizability of the findings. Paid advertisements were posted to groups on Facebook, with target recipients representing diverse demographics. Yet, the sample was predominately white, married, and college educated. The Facebook posts advertising the study likely played a role in the homogenous sample characteristics, as individuals may have subsequently shared the post with friends and family who possessed similar demographic characteristics. In addition, the pregnant women who participated in the study were also a self-selected group who may have possessed a stronger interest in or deterrence to Facebook than the average pregnant woman. Future research would benefit from utilizing additional recruitment methods that are more enticing and accessible to a wider diversity of individuals. In particular, face-to-face and nonsocial media–related recruitment may reach a more diverse group of women. In addition, overall participation and survey completion may have been increased, resulting in a larger sample size, if an incentive had been offered to participants who completed the survey.

Aside from sample characteristics, there were additional limitations to this study which are worthy of note. As previously discussed, the women were not asked *when* they shared the pregnancy-related information, nor were they asked *where* they shared the information. Given that recruitment advertisements were posted to group pages that focused on pregnancy, one can assume that some portion of the sample participated in Facebook groups related to pregnancy. Although the women did indicate who they shared the information with, the categories were not specific enough to ascertain if the information they shared with friends or everyone was shared on group pages or on their own personal Facebook page or if Facebook group members fit within their definition of friends for survey purposes. It would be interesting to investigate if there are differences in terms of frequency, content, and rationale for posts of pregnancy-related information to one's own page versus those to group pages. Such information would be useful to professionals as they seek to incorporate social media within their services for pregnant women.

Finally, the timing of data collection may represent a limitation of the study. Participants were required to reflect upon their use of Facebook during their pregnancy. Recall error may have occurred when reporting how often, what types, and reasons for sharing pregnancy-related information on Facebook. In fact, Moore and McElroy encouraged researchers to use "actual Facebook data where possible and rely on survey data for information that cannot be obtained objectively" [53]. In addition to analyzing actual Facebook posts, an alternative method would involve requiring participants to record their daily use of Facebook for sharing pregnancy-related information through a Web-based portal or mobile app. These methods would overcome the challenge of participants accurately reflecting on their posting of pregnancy-related information on Facebook.

Conclusions

The results of this study fill a gap in our knowledge about pregnant women's use of Facebook to share information about their pregnancy, as well as the role of prenatal attachment in such sharing. The findings supplement previous research linking prenatal attachment to healthy behaviors and self-care by also showing a link between prenatal attachment and sharing information about one's unborn child and pregnancy via social media. Although one certainly cannot endorse or encourage assessing prenatal attachment through Facebook posts, the results of this study are valuable in terms of the additional insight provided regarding associations between prenatal attachment and maternal behaviors.

Perhaps equally, or more, important to our existing knowledge are the descriptive findings of this study and their implications. The pregnant women in this study used Facebook to share pregnancy-related information for a variety of reasons, demonstrating the use of the social media platform during pregnancy for relational, supportive, and informational purposes. In particular, the findings support the suggestion of Bartholomew et al that "conceptions of new parents' social support networks need to be expanded to include the online environment in addition to family, friends, and community members that new parents may see face to face" [6]. To maximize support networks, particularly for pregnant women who lack proximal support, professionals should be cognizant of the potential value of support garnered through social media. Finally, given that pregnant women in this study sought information and advice through Facebook, combined with the fact that an abundance of inaccurate information is available on the Internet and likely shared on social media, the current research further legitimizes a need for professionals to utilize Facebook and other social media platforms to dispense medically accurate information to pregnant women. In particular, support and information may be delivered by medical professionals through closed Facebook groups for patients to "join" [13], Facebook pages that patients "like" to receive medically accurate information and links to other credible sources [6], and Facebook Live sessions that allow pregnant women to interact with professionals in a "live question-and-answer online forum" [10]. Overall, the findings of this study lend merit to the use of Facebook by antenatal medical professionals and educators who are seeking alternative methods for providing information and fostering support among pregnant women via social media in our technology-driven society.

Conflicts of Interest

None declared.

References

1. McDaniel BT, Coyne SM, Holmes EK. New mothers and media use: associations between blogging, social networking, and maternal well-being. Matern Child Health J 2012 Oct;16(7):1509-1517. [doi: 10.1007/s10995-011-0918-2] [Medline: 22094592]
2. Perrin A. Pew Internet. 2015 Oct 05. Social media usage 2005 - 2015 URL: http://www.pewinternet.org/2015/10/08/social-networking-usage-2005-2015/ [accessed 2018-02-27] [WebCite Cache ID 6xYSwVPjD]
3. Statistic Brain. Facebook company statistics 2015 URL: https://www.statisticbrain.com/facebook-statistics/ [WebCite Cache ID 6xYT63YSY]
4. Duggan M, Lenhart A, Lampe C, Ellison NB. Pew Internet. 2015 Jul 16. Parents and social media URL: http://www.pewinternet.org/2015/07/16/parents-and-social-media/ [accessed 2018-02-27] [WebCite Cache ID 6xYTEeNPF]
5. Hicks S, Brown A. Higher Facebook use predicts greater body image dissatisfaction during pregnancy: the role of self-comparison. Midwifery 2016;40:132-140. [doi: 10.1016/j.midw.2016.06.018] [Medline: 27428109]
6. Bartholomew MK, Schoppe-Sullivan SJ, Glassman M, Kamp Dush CM, Sullivan JM. New parents' Facebook use at the transition to parenthood. Fam Relat 2012;61(3):455-469 [FREE Full text] [doi: 10.1111/j.1741-3729.2012.00708.x] [Medline: 23671354]
7. Daniels M, Wedler JA. Enhancing childbirth education through technology. Int J Childbirth Educ 2015;30:28-32.
8. Tomfohrde OJ, Reinke JS. Breastfeeding mothers: use of technology while breastfeeding. Comput Human Behav 2016;64:556-561 [FREE Full text]
9. Song FW, West JE, Lundy L, Dahmen NS. Women, pregnancy, and health information online. Gend Soc 2012;26(5):773-798. [doi: 10.1177/0891243212446336]
10. Lupton D. The use and value of digital media for information about pregnancy and early motherhood: a focus group study. BMC Pregnancy Childbirth 2016;16(1):171 [FREE Full text] [doi: 10.1186/s12884-016-0971-3] [Medline: 27435182]
11. Weston C, Anderson JL. Internet use in pregnancy. Br J Midwifery 2014;22:488-493 [FREE Full text] [doi: 10.12968/bjom.2014.22.7.488]
12. Asiodu IV, Waters CM, Dailey DE, Lee KA, Lyndon A. Breastfeeding and use of social media among first-time African American mothers. J Obstet Gynecol Neonatal Nurs 2015;44(2):268-278 [FREE Full text] [doi: 10.1111/1552-6909.12552] [Medline: 25712127]
13. Haslam DM, Tee A, Baker S. The use of social media as a mechanism of social support in parents. J Child Fam Stud 2017;26(7):2026-2037. [doi: 10.1007/s10826-017-0716-6]
14. Thoren EM, Metze B, Bührer C, Garten L. Online support for parents of preterm infants: a qualitative and content analysis of Facebook 'preemie' groups. Arch Dis Child Fetal Neonatal Ed 2013;98(6):F534-F538. [doi: 10.1136/archdischild-2012-303572] [Medline: 23759520]
15. Hampton K, Rainie L, Lu W, Shin H, Purcell K. Pew Internet. 2015 Jan 15. Social media and the cost of caring URL: http://www.pewinternet.org/2015/01/15/social-media-and-stress/ [WebCite Cache ID 6xYTbP6Av]
16. Chalklen C, Anderson H. Mothering on Facebook: exploring the privacy/openness paradox. Soc Med Soc 2017;3(2):1-10 [FREE Full text] [doi: 10.1177/2056305117707187]
17. Mitchell L. Baby's First Picture: Ultrasound and the Politics of Fetal Subjects. Toronto, Canada: University of Toronto Press; 2001:9780802083494.
18. Roberts J, Griffiths FE, Verran A, Ayre C. Why do women seek ultrasound scans from commercial providers during pregnancy? Sociol Health Illn 2015;37:594-609 [FREE Full text] [Medline: 26094706]
19. Saetnan AR. Thirteen women's narratives of pregnancy, ultrasound, and self. In: Saetnan AR, Oudshoorn N, Kirejczyk MS, editors. Bodies of Technology: Women's Involvement in Prenatal Medicine. Columbus, OH: Ohio State University Press; 2000:331-354.
20. Johnson SA. Maternal devices, social media and the self-management of pregnancy, mothering and child health. Societies (Basel) 2014 Jun 13;4(2):330-350. [doi: 10.3390/soc4020330]
21. Wallwiener S, Muller M, Doster A, Laserer W, Brucker SY, Wallwiener CW, et al. Pregnancy eHealth and mHealth: user proportions and characteristics of pregnant women using web-based information sources: a cross-sectional study. Arch Gynecol Obstet 2016;5:937-944 [FREE Full text]
22. Baker S, Sanders MR, Morawska A. Who uses online parenting support? A cross-sectional survey exploring Australian parents' internet use for parenting. J Child Fam Stud 2016;26(3):916-927 [FREE Full text]
23. Cranley MS. Development of a tool for the measurement of maternal attachment during pregnancy. Nurs Res 1981;30(5):281-284. [doi: 10.1097/00006199-198109000-00008]
24. Muller ME. Development of a prenatal attachment inventory. West J Nurs Res 1993;15:199-215. [doi: 10.1177/019394599301500205]
25. Doan H, Zimmerman A. Conceptualizing prenatal attachment: toward a multidimensional view. J Prenat Perinat Psychol Health 2003;18:131-148.
26. Doan H, Zimmerman A. Prenatal attachment: a developmental model. Int J Prenat Perinat Psychol Med 2008;20:20-28 [FREE Full text]

27. Krisjanous J, Richard JE, Gazley A. The perfect little bump: does the media portrayal of pregnant celebrities influence prenatal attachment? Psychol Mark 2014;31(9):758-773. [doi: 10.1002/mar.20732]
28. Condon JT, Corkindale C. The correlates of antenatal attachment in pregnant women. Br J Med Psychol 1997;70:359-372. [doi: 10.1111/j.2044-8341.1997.tb01912.x]
29. Andrek A, Kekecs Z, Hadhazi E, Boukydis Z, Varga K. Re-evaluation of the psychometric properties of the Maternal-Fetal Attachment Scale in a Hungarian sample. J Obstet Gynecol Neonatal Nurs 2016 Sep;45(5):e15-e25. [doi: 10.1016/j.jogn.2016.05.005] [Medline: 27486088]
30. van Bussel JC, Spitz B, Demyttenaere K. Three self-report questionnaires of the early mother-to-infant bond: reliability and validity of the Dutch version of the MPAS, PBQ and MIBS. Arch Womens Ment Health 2010;13(5):373-384. [doi: 10.1007/s00737-009-0140-z] [Medline: 20127128]
31. de Cock ES, Henrichs J, Vreeswijk CM, Maas AJ, Rijk CH, van Bakel H. Continuous feelings of love? The parental bond from pregnancy to toddlerhood. J Fam Psychol 2016;30(1):125-134. [doi: 10.1037/fam0000138] [Medline: 26280095]
32. Nichols MR, Roux GM, Harris NR. Primigravid and multigravid women: prenatal perspectives. J Perinat Educ 2007;16(2):21-32 [FREE Full text] [doi: 10.1624/105812407X192019] [Medline: 18311335]
33. Alhusen JL, Gross D, Hayat MJ, Woods AB, Sharps PW. The influence of maternal-fetal attachment and health practices on neonatal outcomes in low-income, urban women. Res Nurs Health 2012;35(2):112-120 [FREE Full text] [doi: 10.1002/nur.21464] [Medline: 22262085]
34. Barone L, Lionetti F, Dellagiulia A. Maternal-fetal attachment and its correlates in a sample of Italian women: a study using the Prenatal Attachment Inventory. J Reprod Infant Psychol 2014;32(3):230-239. [doi: 10.1080/02646838.2014.883596]
35. Boukydis CF, Treadwell MC, Delaney-Black V, Boyes K, King M, Robinson T, et al. Women's responses to ultrasound examinations during routine screens in an obstetric clinic. J Ultrasound Med 2006;25:721-728. [doi: 10.7863/jum.2006.25.6.721]
36. De Jong-Pleij EA, Ribbert LS, Pistorius LR, Tromp E, Mulder EJ, Bilardo CM. Three-dimensional ultrasound and maternal bonding, a third trimester study and a review. Prenat Diagn 2013;33(1):81-88. [doi: 10.1002/pd.4013] [Medline: 23169046]
37. Georgsson Ohman S, Waldenstrom U. Effect of first-trimester ultrasound screening for Down syndrome on maternal-fetal attachment: a randomized controlled trial. Sex Reprod Healthc 2010;1:85-90. [doi: 10.1016/j.srhc.2010.05.00]
38. Sedgmen B, McMahon C, Cairns D, Benzie RJ, Woodfield RL. The impact of two-dimensional versus three-dimensional ultrasound exposure on maternal-fetal attachment and maternal health behavior in pregnancy. Ultrasound Obstet Gynecol 2006;27(3):245-251 [FREE Full text] [doi: 10.1002/uog.2703] [Medline: 16482614]
39. Center for Disease Control and Prevention. Tobacco use and pregnancy 2016 URL: https://www.cdc.gov/reproductivehealth/maternalinfanthealth/tobaccousepregnancy/index.htm [WebCite Cache ID 6xYV8VLNB]
40. Magee SR, Bublitz MH, Orazine C, Brush B, Salisbury A, Niaura R, et al. The relationship between maternal-fetal attachment and cigarette smoking over pregnancy. Matern Child Health J 2014 May;18(4):1017-1022 [FREE Full text] [doi: 10.1007/s10995-013-1330-x] [Medline: 23892790]
41. Massey SH, Bublitz MH, Magee SR, Salisbury A, Niaura RS, Wakschlag LS, et al. Maternal-fetal attachment differentiates patterns of prenatal smoking and exposure. Addict Behav 2015;45:51-56 [FREE Full text] [doi: 10.1016/j.addbeh.2015.01.028] [Medline: 25644587]
42. Ross EJ. Maternal fetal attachment and engagement with antenatal advice. Br J Midwifery 2012;20(8):566-575. [doi: 10.12968/bjom.2012.20.8.566]
43. Maddahi MS, Dolatian M, Khoramabadi M, Talebi A. Correlation of maternal-fetal attachment and health practices during pregnancy with neonatal outcomes. Electron Physician 2016;8(7):2639-2644 [FREE Full text] [doi: 10.19082/2639] [Medline: 27648191]
44. Condon JT. The assessment of antenatal emotional attachment: development of a questionnaire instrument. Br J Med Psychol 1993;66:167-183. [doi: 10.1111/j.2044-8341.1993.tb01739.x]
45. Allen SM, Hawkins AJ. Maternal gatekeeping: mothers' beliefs and behaviors that inhibit greater father involvement in family work. J Marriage Fam 1999;61(1):199-212. [doi: 10.2307/353894]
46. Schoppe-Sullivan SJ, Yavorsky JE, Bartholomew MK, Sullivan JM, Lee MA, Kamp Dush CM, et al. Doing gender online: new mothers' psychological characteristics, Facebook use, and depressive symptoms. Sex Roles 2017;16:276-289. [doi: 10.1007/s11199-016-0640-z]
47. Emery LF, Muise A, Dix EL, Le B. Can you tell that I'm in a relationship? Attachment and relationship visibility on Facebook. Pers Soc Psychol Bull 2014;40(11):1466-1479. [doi: 10.1177/0146167214549944] [Medline: 25231798]
48. Collins NL, Allard LM. Cognitive representations of attachment: the content and function of working models. In: Fletcher GJ, Clark MS, editors. Blackwell Handbook of Social Psychology: Vol. 2. Interpersonal Processes. Oxford, United Kingdom: Blackwell; 2002:60-85.
49. Caci B, Cardaci M, Tabacchi ME, Scrima F. Personality variables as predictors of Facebook usage. Psychol Rep 2014;114(2):528-539. [doi: 10.2466/21.09.PR0.114k23w6] [Medline: 24897905]
50. Ryan T, Xenos S. Who uses Facebook? An investigation into the relationship between the big five, shyness, narcissism, loneliness, and Facebook usage. Comput Human Behav 2011;27(5):1658-1664. [doi: 10.1016/j.chb.2011.02.004]

51. Rothman BK. The Tentative Pregnancy: How Amniocentesis Changes the Experience of Motherhood. New York, NY: W. W. Norton & Company; 1986.
52. Ross EJ. I think it's self-preservation: risk perception and secrecy in early pregnancy. Health Risk Soc 2015;17(5-6):329-348. [doi: 10.1080/13698575.2015.1091922]
53. Moore K, McElroy JC. The influence of personality on Facebook usage, wall postings, and regret. Comput Human Behav 2012;28(1):267-274. [doi: 10.1016/j.chb.2011.09.009]

Abbreviations

H: hypothesis
RQ: research question

Edited by G Eysenbach; submitted 24.03.17; peer-reviewed by S Walker, M Ashford; comments to author 04.05.17; revised version received 31.07.17; accepted 04.11.17; published 22.03.18.

Please cite as:
Harpel T
Pregnant Women Sharing Pregnancy-Related Information on Facebook: Web-Based Survey Study
J Med Internet Res 2018;20(3):e115
URL: http://www.jmir.org/2018/3/e115/
doi:10.2196/jmir.7753
PMID:29567636

Review

Web 2.0 Tools in the Prevention of Curable Sexually Transmitted Diseases: Scoping Review

María Sanz-Lorente[1*], RN, MS Clin Nutr; Carmina Wanden-Berghe[2*], MD, PhD; Ramón Castejón-Bolea[1*], MD, PhD; Javier Sanz-Valero[1*], MPH, PhD

[1]Department of Public Health & History of Science, School of Medicine, University Miguel Hernandez of Elche, Alicante, Spain

[2]Foundation for the Promotion of Health and Biomedical Research from the Valencian Community, University General Hospital of Alicante, Alicante, Spain

[*]all authors contributed equally

Corresponding Author:
Javier Sanz-Valero, MPH, PhD
Department of Public Health & History of Science
School of Medicine
University Miguel Hernandez of Elche
Campus Sant Joan d'Alacant
Alicante,
Spain
Phone: 34 666 840 787
Email: jsanz@umh.es

Abstract

Background: The internet is now the primary source of information that young people use to get information on issues related to sex, contraception, and sexually transmitted infections.

Objective: The goal of the research was to review the scientific literature related to the use of Web 2.0 tools as opposed to other strategies in the prevention of curable sexually transmitted diseases (STDs).

Methods: A scoping review was performed on the documentation indexed in the bibliographic databases MEDLINE, Cochrane Library, Scopus, Cumulative Index to Nursing and Allied Health Literature, Web of Science, Literatura Latinoamericana y del Caribe en Ciencias de la Salud, PsycINFO, Educational Resources Information Center, the databases of Centro Superior de Investigaciones Científicas in Spain, and the Índice Bibliográfico Español de Ciencias de la Salud from the first available date according to the characteristics of each database until April 2017. The equation search was realized by means of the using of descriptors together with the consultation of the fields of title register and summary with free terms. Bibliographies of the selected papers were searched for additional articles.

Results: A total of 627 references were retrieved, of which 6 papers were selected after applying the inclusion and exclusion criteria. The STDs studied were chlamydia, gonorrhea, and syphilis. The Web 2.0 tools used were Facebook, Twitter, Instagram, and YouTube. The 6 papers used Web 2.0 in the promotion of STD detection.

Conclusions: Web 2.0 tools have demonstrated a positive effect on the promotion of prevention strategies for STDs and can help attract and link youth to campaigns related to sexual health. These tools can be combined with other interventions. In any case, Web 2.0 and especially Facebook have all the potential to become essential instruments for public health.

(J Med Internet Res 2018;20(3):e113) doi:10.2196/jmir.8871

KEYWORDS

sexually transmitted diseases, bacterial; internet; webcasts; social media

Introduction

General measures of health promotion and education are fundamental in the prevention of sexually transmitted diseases (STDs), especially favorable strategies for safe sex. Health education on the symptoms of these diseases, methods of transmission, prevention, diagnosis, and treatment are main measures of control [1].

STDs have profound effects on sexual and reproductive health worldwide and are among the 5 major categories for which

adults seek health care. Every day more than 1 million people contract a sexually transmitted infection. It is estimated that annually, about 357 million people contract any of 4 curable STDs: chlamydia (131 million), gonorrhea (78 million), syphilis (5.6 million), or trichomoniasis (143 million) [2].

In the 21st century, the epidemiological evolution of STDs cannot be understood without taking into account factors such as globalization, migration, and the information and communication technologies (ICTs) that have led to new approaches in the study of their transmission and prevention [3]. As a result, sex education must be addressed from all facets of possible action making sure that the information is complete. In this last aspect, neither school nor the family seems to be sufficient [4]. The usual methods used in education for sexuality, such as workshops and presentations, among others, play an important role in the transfer of information; however, today's young adult has different interests. In this regard, Jimenez [5] states that "young people have a constant provision to the use and management, contact and utilization of technological gadgets; taking these to make them partakers of their life in whatever place and moment required."

As for health education, it is necessary that the content system and messages related to the prevention reach young people in the most informal and entertaining way, for which ICTs would be very useful. It is well known that the dawn of Web 2.0 resources has provoked a substantive change in the communication of knowledge, favoring its disclosure by enabling the expansion and permeability of knowledge at a very low cost. Web 2.0 has shown its integration in today's information society and, far from dwindling, increasingly has more initiatives that enhance it, subsequently contributing to the diffusion of the contents about health [6].

In Spain, in a survey conducted by Doctoralia internet in 2016 [7], young people between 18 and 24 years old were the most prone to self-medication (41%), and 7% of them have made a mistake by choosing a medication or searching for a solution for their health problem on the internet. At the time, 69% of this group sought information on the internet after being diagnosed with a condition. One-quarter (26%) confessed to having lied or hidden information from their doctor, doing so because they didn't want to reveal some aspect of their intimacy, they felt they had done something wrong to their health, or they felt shame at the time of appointment or consultation, especially with the urologist (21%).

Despite easy access to health professionals in specialized units, many young girls get their information from friends and on the Web; the internet is now the primary source of information that young people use to get information on issues related to sex, contraception, and sexually transmitted infections [8]. The vast majority of teenagers search on the internet because of its anonymity without taking into account that not everything they find will be true [9].

Information can improve people's ability to recognize the symptoms of STDs, increasing the chances that they will request medical attention or encourage their partners to do so [2].

In this context, the objective of this systematic review was to evaluate studies that use the Web 2.0 in contrast with other strategies to prevent curable STDs.

Methods

Data were obtained from the following bibliographic databases in the field of health science: MEDLINE (via PubMed), Cochrane Library, Scopus, Cumulative Index to Nursing and Allied Health Literature, Web of Science, *Literatura Latinoamericana y del Caribe en Ciencias de la Salud*, PsycINFO, Educational Resources Information Center, the bibliographic databases of the *Centro Superior de Investigaciones Científicas* of Spain, and the *Índice Bibliográfico Español de Ciencias de la Salud*.

Information Processing

Search terms were chosen from the thesaurus developed by the US National Library of Medicine (Medical Subject Headings [MeSH] and title/abstract), and the final search syntax was shaped by the Boolean intersection of 2 equations (equation 1 AND equation 2):

- Equation 1: ("internet"[MeSH] OR "Social Media"[MeSH] OR "internet"[Title/Abstract] OR "World Wide Web"[Title/Abstract] OR "WWW"[Title/Abstract] OR "Web"[Title/Abstract] OR "Social Media"[Title/Abstract] OR "Blog"[Title/Abstract] OR "Wikipedia"[Title/Abstract] OR "Wiki"[Title/Abstract] OR "YouTube"[Title/Abstract] OR "Facebook"[Title/Abstract] OR "Twitter"[Title/Abstract])
- Equation 2: ("Sexually Transmitted Diseases, Bacterial"[MeSH] OR "Trichomonas Infections"[MeSH] OR "Bacterial Sexually Transmitted Disease"[Title/Abstract] OR "Sexually Transmitted Diseases, Bacterial"[Title/Abstract] OR "Bacterial STIs"[Title/Abstract] OR "Bacterial STDs"[Title/Abstract] OR "Bacterial Venereal Disease"[Title/Abstract] OR "Bacterial Sexually Transmitted Infection"[Title/Abstract] OR "Venereal Diseases, Bacterial"[Title/Abstract] OR "Chancroid"[Title/Abstract] OR "Lymphogranuloma Venereum"[Title/Abstract] OR "Trachoma"[Title/Abstract] OR "Chlamydia"[Title/Abstract] OR "Chlamydia Infection"[Title/Abstract] OR "Gonorrhea"[Title/Abstract] OR "Neisseria"[Title/Abstract] OR "Granuloma Inguinale"[Title/Abstract] OR "Granuloma Venereum"[Title/Abstract] OR "*Haemophilus ducreyi*"[Title/Abstract] OR "Donovanosis"[Title/Abstract] OR "Syphilis"[Title/Abstract] OR "Treponema"[Title/Abstract] OR "Great Pox"[Title/Abstract] OR "Chancre"[Title/Abstract] OR "*Klebsiella granulomatis*"[Title/Abstract] OR "Calymmatobacterium"[Title/Abstract] OR "*Mycoplasma genitalium*"[Title/Abstract] OR "*Ureaplasma urealyticum*"[Title/Abstract] OR "*Trichomonas vaginalis*"[Title/Abstract] OR "Trichomonas Infection"[Title/Abstract] OR "Trichomonas vaginitis"[Title/Abstract])

The final search equation was developed for use in the database MEDLINE, via PubMed, using the filters: “Humans” and “Comparative Study” or “Evaluation Studies.”

This strategy was adapted to the characteristics of each of the rest of the databases consulted. The search was carried out from the first available date according to the characteristics of each database until April 2017 and was completed with the consideration of the bibliographic listing of the items that were selected.

Final Selection of Papers

Papers were selected that met the following criteria (criteria of inclusion): comply with the objectives of the search, published in journals reviewed by peers, and written in English, Spanish, Portuguese, French, or German. Papers that did not present results about the advantages of Web 2.0 in relation to other strategies for the prevention of curable STDs were excluded.

The selection of the relevant papers was performed independently by 2 authors (MSL and JSV). For inclusion of the studies, it was established that the valuation of the concordance between these authors (kappa index) must be greater than .80. Provided this condition is fulfilled, possible discrepancies were solved through consultation with the author CWB and subsequent consensus among all the authors [10].

The quality of the selected documents was evaluated using the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines [11], which contain a list of 22 essential items that must be described in the studies. For each selected paper, 1 point was assigned for each present item (not applicable=0). When an item addressed several issues, these were evaluated independently, giving partial value to each one and averaging so that in no case could the value be more than one.

Data Extraction

Control of the correctness of the data was performed using double tables that allowed the detection of deviations and their correction by revising the originals. The semiperiod of Burton-Kebler (the median of age) and Price (percentage of papers less than 5 years old) indices were calculated to determine the relevance of papers. The studies were grouped according to the variables to study in order to systematize and facilitate the understanding of the results, considering the following data: first author of the bibliographic reference and year of publication, type of study, country and age of the participants, curable STD discussed, Web 2.0 tool used in the study, period in which the work was done, intervention carried out, and results obtained.

Results

A total of 627 references were retrieved, and 1 paper was obtained from the bibliographic listings of relevant retrieved papers.

After debugging the duplicates, applying inclusion and exclusion criteria, and consulting the bibliographic lists (see Figure 1), 6 documents [12-17] were selected for review and critical analysis (Multimedia Appendix 1). The calculation of kappa coefficient gave a measure of agreement on the selection of the papers, between evaluators, of .96 (P<.001).

The 6 selected papers presented an obsolescence, according to the Burton-Kebler index, equal to 1 year, with a Price index of 100%. When assessing the quality of papers selected for review using the STROBE questionnaire, scores ranged between 8.33 and 17.00, with a median of 13.51 (Multimedia Appendix 2).

The revised works were 3 evaluation studies [12,15,16] and 3 comparative studies [13,14,17]; 5 developed in the United States [12,14-17] and 1 in New Zealand [13]; all written in English.

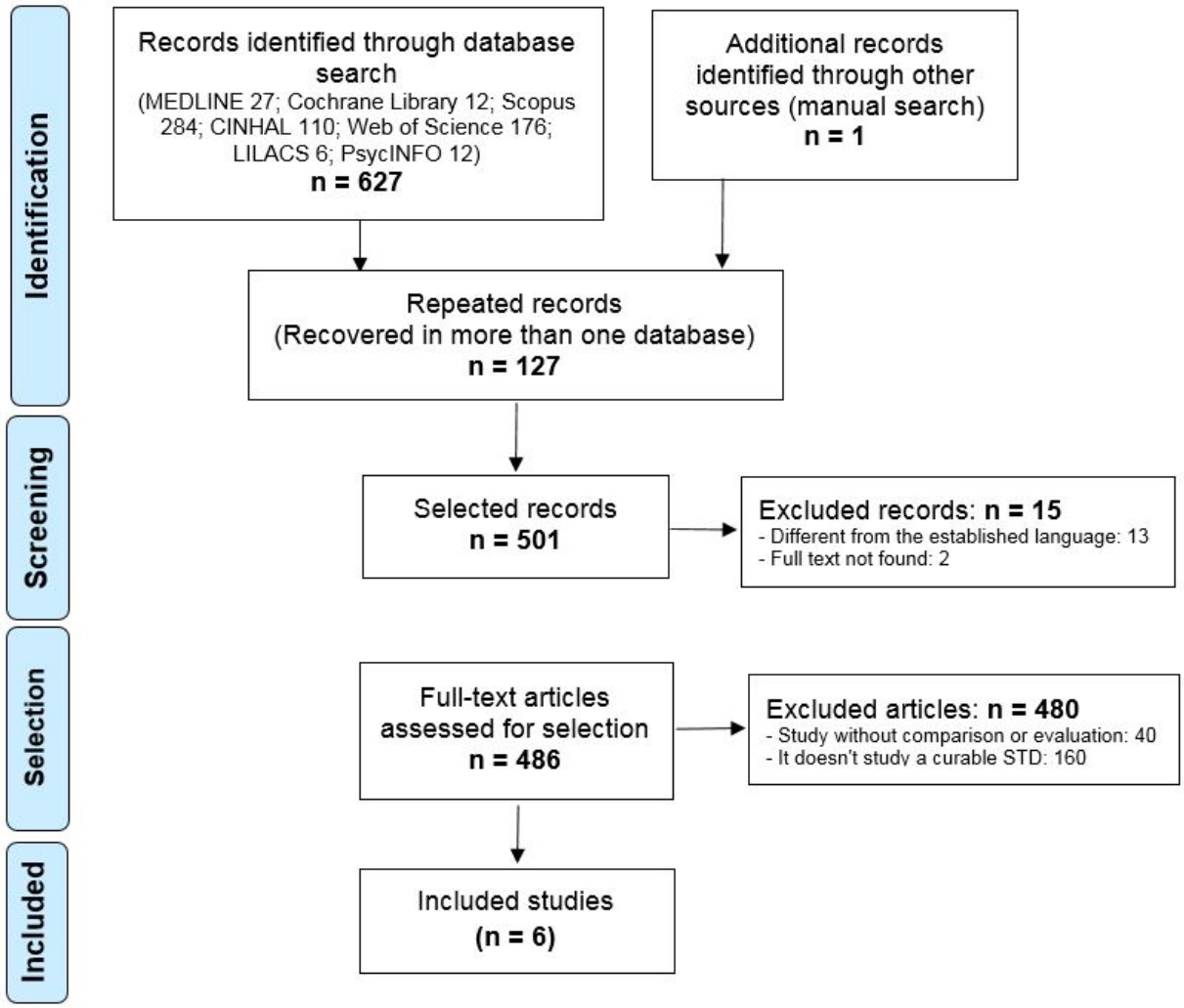
All studies were developed in people aged 25 years or less except Habet et al [14], which included participants up to 35 years old in its second phase. The curable STD targets of these works were chlamydia [12,14,16,17], gonorrhea [12,14,17], syphilis [12,13], and any STD [15].

The longest period of implementation of a promotion about STD testing was the Get Yourself Tested (GYT) campaign [18] through the Division of Sexually Transmitted Disease Prevention, Centers for Disease Control and Prevention (CDC), observed in Friedman et al [15].

Facebook was the Web 2.0 tool used in the 6 papers, although Dowshen et al [12] also used Twitter, Instagram, and YouTube and Friedman et al [15] also used Twitter. The results of user interactions were offered in 3 works: Dowshen et al [12] noted approximately 6000 visits to Facebook and 128 likes, 46 followers on Twitter, 390 YouTube views, and 42 Instagram followers; interaction data in Friedman et al [15] offered 4477 Facebook likes and 1994 Twitter followers; and Bull et al [17] indicated on average 43 visits per week (range of 37 to 101).

The Dowshen et al [12], Coughlan et al [13], and Friedman et al [15] papers determined the usefulness of Web 2.0 tools for increasing awareness in and implementation of screening. Jones et al [16] and Bull et al [17] identified an increase in condom use and positive changes in behavior among the participant population as a result of the promotion campaign. On the contrary, Habel et al [14] did not observe favorable differences in relation to testing, indicating that it would have been a key to the training and collaboration of health care personnel in support of the campaign. The Dowshen et al [12] and Bull et al [17] papers also reported a reduction in positive cases.

Figure 1. Identification and selection of studies.



Discussion

Principal Findings

The results of this review show that Web 2.0 tools can be useful in the prevention of curable STDs. The identification of just 6 publications is not surprising since an exploratory review on the promotion of sexual health through social networks found 51 papers, mostly focusing on HIV infection; no assessment or comparison with traditional methods was performed. In addition, the authors of this work pointed out that most of the efforts to implement ICT in the promotion of STD prevention campaigns had developed in HIV campaigns despite an increase in the incidence of curable STDs [19].

On the other hand, the high number of nonrelevant papers was mainly due to results obtained from the Web of Science and Scopus databases, which do not have descriptor thesauri. Queries are constructed by entering text in title, abstract, and keyword fields. This high documentary noise has been observed in other systematic reviews [20,21].

In the papers selected for review, validity and topicality were verified. The data obtained indicate a lower obsolescence than has been observed in works previously published in the field of health sciences. Moreover, it is evident that the results derived from the age of publication (measured by the median and Price index) is a characteristic of an area of knowledge in full emergence [22].

The fact that the documents included in the review were written in English and came mostly from US institutions was an expected fact in line with the existing bibliometric results [23].

The age of the population included in the reviewed studies coincides with the age group with greater incidence of STDs [24]. Statistics of major international health agencies show that young people are most affected by STDs, and these consequences can affect the rest of their lives. The vast majority declare being sexually active and protect themselves from pregnancy but not from STDs. In general, they show little knowledge of sexual transmission of infections, although they know of the concept. Syphilis is considered a disease of other people. Some knew about gonorrhea but most had not heard of chlamydia and did not perceive themselves to be at risk [25].

Young people say loneliness and abandonment are to blame for the lack of information about their sexuality. Thus, the most frequent source of information is friends, then the internet, traditional media (especially television), parents, and finally

medical consultation. University organizations are rarely mentioned, except for sporadic or very specific initiatives [26].

At the same time, of the curable STDs studied, chlamydia, gonorrhea, and syphilis have the highest incidence and prevalence [27], which is an an adequate representation of these diseases.

The presence of Facebook in all the studies reviewed is logical; this Web 2.0 tool has been placed among the 3 most commonly used in the world and has already shown its potential for health promotion [28,29]. As the CDC indicates, Facebook is a tool of great potential for its use in different prevention programs and health promotions [30].

In recent years, Facebook, YouTube, Twitter, and other Web 2.0 tools have become effective ways to expand the reach, promote commitment, and increase access to messages on health and prevention and treatment of STDs [30-32].

Little attention was given to user interaction. Only half of the works described these data (visits, likes, followers, etc) and none assessed them. These data could have provided interesting results complementary to those that could had been obtained through traditional systems of public health surveillance, as is seen in the recent work of Gittleman et al [33]. The essence of the user interaction with the materials of the Web 2.0 lies in knowing the interest generated together with a wide range of services enabling collaboration and fast exchange of information among users of a community or social network.

Users can communicate with the issuers of material and show appreciation through a symbolic, easily understandable code (I like/dislike) or even by emotional expressions (I love it, I enjoy it, it saddens me, etc) in an agile exchange of information facilitated by the structure and design of the website. Research shows that even low user interaction or passivity is something attractive in the content consumer, probably by the distrust generated by not knowing who is on the other side of the screen or, simply, to avoid being observed (monitored) when they have to register (give personal data) to be able to interact with the Web tool [34].

The revised works focused their intervention on the awareness and prevention of STDs, with particular emphasis on the promotion of screening. However, until now there has not been much evidence about how the information on the Web influences people's health behavior, which is necessary to deepen the study of the usefulness of social networks for the benefit of the promotion of health [35]. Thus, Taggart et al [36] and Hochberg et al [37], in 2 separate systematic reviews on HIV published in 2015, pointed to the need for further research to determine to what extent ICTs can influence the prevention of STDs. This recommendation also appeared in all revised papers [12-17].

Limitations

A limitation of this review could be the low number of selected papers because it is an area of emerging technological application. It has been stated that systematic reviews should be based on studies with design and selection that ensure greater scientific rigor, but in this analysis, all retrieved papers focusing on the studied subject were included.

According to the US Agency for Health Research and Quality, epidemiological designs of the studies selected in this review do not guarantee full validity and reliability of the obtained observations. However, the evidence available is probably the best, given the difficulties of study in this area of research and based on the observations obtained in the different interventions. As a result, while it would have been more interesting to have a specific questionnaire, it was considered appropriate to use the STROBE questionnaire to evaluate the quality of the studies.

Although the real limitations are due to the characteristic of each study *per se*, from these limitations, important lessons in formulating appropriate actions for the development, implementation, and evaluation of future Web 2.0 applications can be extracted.

Conclusions

For all of these reasons, we conclude that Web 2.0 tools have demonstrated a positive effect on the promotion of prevention strategies for STDs and can help attract and link young people to campaigns related to sexual health. These tools can even be combined with other interventions. In any case, Web 2.0 tools, especially Facebook, have all the potential to become key instruments in public health.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Characteristics and main findings of the studies selected for review.

[PDF File (Adobe PDF File), 36KB - jmir_v20i3e113_app1.pdf]

Multimedia Appendix 2

Methodological quality of the studies based on the 22-point assessment from the Strengthening the Reporting of Observational Studies in Epidemiology guideline.

[PDF File (Adobe PDF File), 26KB - jmir_v20i3e113_app2.pdf]

References

1. Centro Nacional de Epidemiología, Instituto de Salud Carlos III. Red Nacional de Vigilancia Epidemiológica. Protocolos de enfermedades de declaración obligatoria. Madrid: Dirección General de Salud Pública, Calidad e Innovación, Ministerio de Sanidad Servicios Sociales e Igualdad; 2013. URL: https://goo.gl/wW97P9 [accessed 2018-03-01] [WebCite Cache ID 6xbEyuoes]
2. Sexually transmitted infections (STIs). Geneva: World Health Organization; 2016. URL: http://www.who.int/mediacentre/factsheets/fs110/es/ [accessed 2018-03-01] [WebCite Cache ID 6xbFOARhN]
3. Nahmias SB, Nahmias D. Society, sex, and STIs: human behavior and the evolution of sexually transmitted diseases and their agents. Ann N Y Acad Sci 2011 Aug;1230:59-73. [doi: 10.1111/j.1749-6632.2011.06079.x] [Medline: 21824166]
4. Sociedad Española de Médicos de Atención Primaria (SEMERGEN). Promoción de la salud y prevención de la enfermedad en la mujer española del siglo XXI. Madrid: SEMERGEN; 2009. URL: http://www.munideporte.com/imagenes/documentacion/ficheros/20100419103253estudio_semergen.pdf [accessed 2018-03-01] [WebCite Cache ID 6xbFUj7nJ]
5. Jiménez E. El nino tecnologico: un perfil educativo. Rev Investig 2006;60:119-131.
6. Sanz-Lorente M, Guardiola-Wanden-Berghe R, Wanden-Berghe C, Sanz-Valero J. Presence and characteristics of nursing terminology in Wikipedia. Rev Enferm 2013 Oct;36(10):22-26. [Medline: 24354129]
7. NetQuest Investigaciones. Los españoles y la medicina 2.0. Barcelona: Doctoralia Internet SL; 2016. URL: http://insights.doctoralia.es/los-espanoles-y-la-medicina-2-0/ [accessed 2018-03-01] [WebCite Cache ID 6xbFh9p4S]
8. Úbeda A. Internet primera fuente de información sobre sexualidad de las adolescentes. : Hospital Universitario Quirón; 2014 Presented at: Symposium Internacional: Actualización en ginecología general; Nov 6-7, 2014; Barcelona p. 41.
9. Suárez M. Internet primera fuente de información sobre sexualidad de las adolescentes. : Hospital Universitario Quirón; 2014 Presented at: Symposium Internacional: Actualización en ginecología general; Nov 6-7, 2014; Barcelona p. 41.
10. Wanden-Berghe C, Sanz-Valero J. Systematic reviews in nutrition: standardized methodology. Br J Nutr 2012 Jun;107 Suppl 2:S3-S7. [doi: 10.1017/S0007114512001432] [Medline: 22591902]
11. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. Declaracion de la Iniciativa STROBE (Strengthening the Reporting of Observational studies in Epidemiology): directrices para la comunicacion de estudios observacionales. Gaceta Sanitaria 2008 Mar;22(2):144-150. [doi: 10.1157/13119325]
12. Dowshen N, Lee S, Matty LB, Castillo M, Mollen C. IknowUshould2: feasibility of a youth-driven social media campaign to promote STI and HIV testing among adolescents in Philadelphia. AIDS Behav 2015 Jun;19 Suppl 2:106-111 [FREE Full text] [doi: 10.1007/s10461-014-0991-9] [Medline: 25563502]
13. Coughlan E, Young H, Parkes C, Coshall M, Dickson N, Psutka R, et al. A novel response to an outbreak of infectious syphilis in Christchurch, New Zealand. Sex Health 2015 Apr;12(2):170-173. [doi: 10.1071/SH14140] [Medline: 25493444]
14. Habel MA, Scheinmann R, Verdesoto E, Gaydos C, Bertisch M, Chiasson MA. Exploring pharmacy and home-based sexually transmissible infection testing. Sex Health 2015 Nov;12(6):472-479 [FREE Full text] [doi: 10.1071/SH15031] [Medline: 26409484]
15. Friedman AL, Brookmeyer KA, Kachur RE, Ford J, Hogben M, Habel MA, et al. An assessment of the GYT: Get Yourself Tested campaign: an integrated approach to sexually transmitted disease prevention communication. Sex Transm Dis 2014 Mar;41(3):151-157. [doi: 10.1097/OLQ.0000000000000100] [Medline: 24521718]
16. Jones K, Baldwin KA, Lewis PR. The potential influence of a social media intervention on risky sexual behavior and Chlamydia incidence. J Community Health Nurs 2012;29(2):106-120. [doi: 10.1080/07370016.2012.670579] [Medline: 22536914]
17. 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]
18. Centers for Disease Control and Prevention. GYT: Get Yourself Tested campaign—STD awareness resources web page. URL: https://npin.cdc.gov/stdawareness/gyt.aspx [accessed 2018-03-01] [WebCite Cache ID 6xbHq13lB]
19. Gabarron E, Wynn R. Use of social media for sexual health promotion: a scoping review. Glob Health Action 2016;9:32193 [FREE Full text] [Medline: 27649758]
20. Domingo-Pueyo A, Sanz-Valero J, Wanden-Berghe C. [Effects of occupational exposure to chromium and its compounds: a systematic review]. Arch Prev Riesgos Labor 2014;17(3):142-153. [doi: 10.12961/aprl.2014.17.3.03] [Medline: 25022532]
21. Domingo-Pueyo A, Sanz-Valero J, Wanden-Berghe C. Disorders induced by direct occupational exposure to noise: systematic review. Noise Health 2016;18(84):229-239 [FREE Full text] [doi: 10.4103/1463-1741.192479] [Medline: 27762251]
22. Martínez GJ, Sanz-Valero J, Wanden-Berghe C. Information systems in clinical pharmacy applied to parenteral nutrition management and traceability: a systematic review. Farm Hosp 2017 Jan 01;41(n01):89-104 [FREE Full text] [doi: 10.7399/fh.2017.41.1.10610] [Medline: 28045654]
23. Franco-López A, Sanz VJ, Culebras J. To publish in Spanish or in any non English language, negative for impact factor and citations. JONNPR 2016;1(2):65-70. [doi: 10.19230/jonnpr.2016.1.2.1005]
24. Global strategy for the prevention and control of sexually transmitted infections. Geneva: World Health Organization; 2007. URL: http://apps.who.int/iris/bitstream/10665/43773/1/9789243563473_spa.pdf [accessed 2018-03-01] [WebCite Cache ID 6xbJUx4CC]

25. Grupo de trabajo sobre ITS. Infecciones de transmisión sexual: Diagnóstico, tratamiento, prevención y control. Madrid: Ministerios de Sanidad, Política Social e Igualdad; 2011. URL: https://goo.gl/y2sEvL [accessed 2018-03-01] [WebCite Cache ID 6xbJYEJtf]
26. Montes de Oca A. Estudiantes universitarios frente a las infecciones de transmision sexual: Un estudio de audiencia desde la comunicacion para la salud. Corresp Análisis 2013;3:367-381.
27. Global incidence and prevalence of selected curable sexually transmitted infections—2008. Geneva: World Health Organization; 2012. URL: http://apps.who.int/iris/bitstream/10665/75181/1/9789241503839_eng.pdf?ua=1 [accessed 2018-03-01] [WebCite Cache ID 6xbJeIEoQ]
28. Alexa—the top 500 sites on the Web. Seattle: Amazon Company URL: https://www.alexa.com/topsites [accessed 2018-03-01] [WebCite Cache ID 6xbJjDHq7]
29. Zhang Y, He D, Sang Y. Facebook as a platform for health information and communication: a case study of a diabetes group. J Med Syst 2013 Jun;37(3):9942. [doi: 10.1007/s10916-013-9942-7] [Medline: 23588823]
30. The health communicator's social media toolkit. Atlanta: Centers for Disease Control and Prevention; 2011. URL: https://www.cdc.gov/healthcommunication/ToolsTemplates/SocialMediaToolkit_BM.pdf [accessed 2018-03-01] [WebCite Cache ID 6xbJr2czf]
31. Hunter P, Oyervides O, Grande KM, Prater D, Vann V, Reitl I, et al. Facebook-augmented partner notification in a cluster of syphilis cases in Milwaukee. Public Health Rep 2014;129 Suppl 1:43-49 [FREE Full text] [doi: 10.1177/00333549141291S107] [Medline: 24385648]
32. Syred J, Naidoo C, Woodhall SC, Baraitser P. Would you tell everyone this? Facebook conversations as health promotion interventions. J Med Internet Res 2014;16(4):e108 [FREE Full text] [doi: 10.2196/jmir.3231] [Medline: 24727742]
33. Gittelman S, Lange V, Gotway CCA, Okoro CA, Lieb E, Dhingra SS, et al. A new source of data for public health surveillance: Facebook likes. J Med Internet Res 2015;17(4):e98 [FREE Full text] [doi: 10.2196/jmir.3970] [Medline: 25895907]
34. Oller-Arlandis V. Calidad y pertinencia de la información existente en la Web 2.0 sobre medio ambiente. Sant Joan d'Alacant: Miguel Hernández University; 2016.
35. Maher CA, Lewis LK, Ferrar K, Marshall S, De Bourdeaudhuij I, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res 2014;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
36. 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 02;17(11):e248 [FREE Full text] [doi: 10.2196/jmir.4387] [Medline: 26525289]
37. Hochberg CH, Berringer K, Schneider JA. Next-generation methods for HIV partner services: a systematic review. Sex Transm Dis 2015 Sep;42(9):533-539 [FREE Full text] [doi: 10.1097/OLQ.0000000000000335] [Medline: 26267881]

Abbreviations

CDC: Centers for Disease Control and Prevention
ICT: information and communication technology
MeSH: Medical Subject Headings
STD: sexually transmitted disease
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by G Eysenbach; submitted 30.08.17; peer-reviewed by A Franco-Perez, L Alvarez-Sabucedo, KA Nguyen; comments to author 18.10.17; revised version received 11.11.17; accepted 13.11.17; published 22.03.18.

Please cite as:
Sanz-Lorente M, Wanden-Berghe C, Castejón-Bolea R, Sanz-Valero J
Web 2.0 Tools in the Prevention of Curable Sexually Transmitted Diseases: Scoping Review
J Med Internet Res 2018;20(3):e113
URL: http://www.jmir.org/2018/3/e113/
doi:10.2196/jmir.8871
PMID:29567633

Review

Use of Social Network Sites for Communication Among Health Professionals: Systematic Review

Windy SY Chan[1,2*], BPharm, MClinPharm; Angela YM Leung[3*], RN, BN, MHA, PhD

[1]School of Health Sciences, Caritas Institute of Higher Education, New Territories, China (Hong Kong)

[2]Faculty of Health and Social Sciences, The Hong Kong Polytechnic University, Hung Hom, Kowloon, China (Hong Kong)

[3]Centre for Gerontological Nursing, School of Nursing, The Hong Kong Polytechnic University, Hung Hom, Kowloon, China (Hong Kong)

[*]all authors contributed equally

Corresponding Author:
Angela YM Leung, RN, BN, MHA, PhD
Centre for Gerontological Nursing
School of Nursing
The Hong Kong Polytechnic University
Room GH528, 5th Floor, Core G,
The Hong Kong Polytechnic University
Hung Hom, Kowloon,
China (Hong Kong)
Phone: 852 27665587
Fax: 852 23649663
Email: angela.ym.leung@polyu.edu.hk

Abstract

Background: Although much research has been done investigating the roles of social network sites (SNSs) in linking patients and health professionals, there is a lack of information about their uses, benefits, and limitations in connecting health professions only for professional communication.

Objective: This review aimed to examine the utilization of SNSs for communication among health professionals in (1) frontline clinical practice, (2) professional networks, and (3) education and training to identify areas for future health communication research.

Methods: This review followed the Preferred Reporting Items for Systematic Reviews and Meta-analyses guidelines. A systematic search of the literature published in the last 10 years (January 1, 2007, to March 1, 2017) was performed in March 2017, using the following electronic databases: MEDLINE via OvidSP, EMBASE, CINAHL Complete, and InfoSci-Journals. The searches were conducted using the following defined search terms: “social media” OR “social network” OR “social network site” OR “Facebook” OR “Twitter” OR “Linkedin” OR “Instagram” OR “Weibo” OR “Whatsapp” OR “Telegram” OR “WeChat” AND “health” OR “health profession.”

Results: Of the 6977 papers retrieved, a total of 33 studies were included in this review. They were exploratory in nature, and the majority used surveys (n=25) and interviews (n=6). All retrieved studies stated that SNSs enhanced effective communication and information sharing. SNSs were used for supporting delivering of clinical services, making referrals, and sharing information. They were beneficial to network building and professional collaboration. SNSs were novel tools to enhance educational interactions among peers, students, instructors, and preceptors. The application of SNSs came with restraints in technical knowledge, concerns on data protection, privacy and liability, issues in professionalism, and data protection.

Conclusions: SNSs provide platforms facilitating efficient communication, interactions, and connections among health professionals in frontline clinical practice, professional networks, education, and training with limitations identified as technical knowledge, professionalism, and risks of data protection. The evolving use of SNSs necessitates robust research to explore the full potential and the relative effectiveness of SNSs in professional communication.

(J Med Internet Res 2018;20(3):e117) doi:10.2196/jmir.8382

KEYWORDS

social networking; social media; health communication; Facebook; WhatsApp; professional network; health professionals

Introduction

Background

Social network sites (SNSs) are Web-based services that allow individuals to construct a profile and build a network of connections with other users within the system [1]. Since their introduction, SNSs have become integrated into the daily practices of millions of users. With the evolving technologies in mobile-based platforms and apps, SNSs are currently constructed as Web 2.0 Internet-based apps [2].

The world's current largest social network, Facebook, has engaged more than 2.01 billion users worldwide [3,4]. Twitter, with more than 330 million of monthly active users, has become essential to scientific conferences, gaining them publicity via sharing real-time proceedings or live-tweeting [5]. SNSs provide platforms for users to share their own content, react, or add comments on the content posted by other users. They help strangers to be connected based on their common interests, activities, identities, or professions. LinkedIn, with more than 530 million members in over 200 countries and territories, focuses on business connections and industry contacts for employers and working professionals. It allows users to enhance their connectedness in their areas of expertise [6]. SNSs differ from traditional broadcast media in supporting networking by information and communication technologies. WhatsApp Messenger brings free, cross-platform communication beyond text-only messages to more than 1 billion people in over 180 countries [7].

Availability and preferences of SNSs vary across countries. Facebook is the top worldwide yet, in some countries, such as Indonesia, Instagram has taken its place, and some African territories prefer LinkedIn [8]. In China, where some SNSs are not available, QZone is the top social network. VKontakte and Odnoklassniki, which are both controlled by Russia's Mail.Ru group, have also gained ground in Russian territories [8].

SNSs are widely used in health communication and research [9] and provide platforms to the public to access health information and to seek support if needed. A new dimension to health care was created to enable the public, patients, and health professionals to communicate about health issues and to give them the possibility of improving health outcomes [10]. In a meta-analysis, SNS interventions were found to be effective in changing health behavior–related outcomes in which the predominant health domain was fitness related (eg, weight loss and physical activity) [11]. Emerging evidence support using SNSs among health professionals to develop virtual communities for sharing domain knowledge [12].

Objective

Most current literature reviews have focused on the roles of SNSs in linking patients and health professionals [9,10,13].Nevertheless, there is a lack of information about the uses, benefits, and limitations of SNSs in connecting health professions only (excluding the involvement of patients). This systematic review aims to examine the utilization of SNSs for communication among health professionals in (1) frontline clinical practice, (2) professional networks, and (3) education and training to identify important areas for health communication research in the future. In the context of this review, frontline clinical practice refers to the delivery and operation of health services; professional networks refer to the interactions and relationships of a professional nature rather than personal interactions; and education and training are meant to be the training of students and professional development in the health care field.

Methods

Search Strategy

This review followed the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines [14]. A systematic search of the literature published in the last 10 years (January 1, 2007, to March 1, 2017) was performed in March 2017, using the following electronic databases: MEDLINE via OvidSP, EMBASE, CINAHL Complete, and InfoSci Journals.

SNSs itself has not been defined as a medical subject headings (MeSH) to optimize retrieval of relevant papers, "social media" (a MeSH term) is used in the search because SNSs are considered as a subset of social media [9]. As the number of SNSs being used rises continuously, the search terms were limited to the top most frequently used ones [3]. The searches were performed using the following search terms: "social media" (a MeSH term) OR "social network" OR "social network site" OR "Facebook" OR "Twitter" OR "Linkedin" OR "Instagram" OR "Weibo" OR "Whatsapp" OR "Telegram" OR "WeChat" AND "health" (a MeSH term) OR "health profession" (a MeSH term).

Initial screening of the studies, based on the information contained in the titles and abstracts, was undertaken independently by 2 reviewers. If a decision on inclusion or exclusion could not be reached, the full text was retrieved. The full texts of the shortlisted papers were then assessed independently by 2 reviewers. The reference lists of relevant papers were also screened for eligible papers. The reviewers met to discuss studies for inclusion and to reach consensus. If there was a discrepancy, a third reviewer was consulted.

Study Inclusion and Exclusion Criteria

This review included all study designs to identify the best evidence available to address the research objective. Studies were included in this review if they (1) focused primarily on communication interactions between and among health professionals about health issues using SNSs and (2) studied the uses, benefits, or limitations of SNSs.

Studies were excluded from this review if they (1) were not in English, (2) were reviews, reports, abstracts only, letters, or commentaries, (3) focused primarily on the communication between public or patients and health professionals, or for personal uses, (4) described the use of SNSs primarily with a marketing or advertising focus, (5) studied non-SNS types of social media (eg, websites, short message service, emails, hospital information systems, and electronic health record systems), or (6) were not available as full text in the final search.

Data Extraction, Synthesis, and Evaluation

A computer-based form was created for data extraction. The data collected included first author, year, country, study type, number of participants, health profession(s) involved, type(s) of SNSs, functions of the SNSs (eg, for education, data sharing, continuous professional development), controls and their characteristics (if applicable), and primary outcome measures (and secondary outcome measures if they were highly relevant). The Critical Appraisal Skills Programme (CASP) appraisal tools were used to evaluate the quality of the reviewed studies. They can be used to critically appraise the evidence of a wide variety of settings and designs (eg, qualitative studies or studies using mixed methods). Each CASP tool consists of 3 sections, and each section is designed to assess different domains of a primary study (the internal validity of the instruments used in the primary study, the results, and the relevance of the findings to practice) [15,16]. The Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies developed by the National Heart, Lung, and Blood Institute was used to evaluate the quality of the quantitative studies [17]. Two reviewers assessed the quality of the included studies independently. If necessary, a third reviewer was involved in settling disagreements.

Ethics Approval and Consent to Participate

This was a systematic review with no data collected from human subjects. Ethical approval was not needed.

Results

Findings

Figure 1 shows the searching process and how the studies were included in this review. The literature search retrieved 6977 papers. Their titles and abstracts were screened, and those that did not meet the inclusion criteria were removed. Duplicated titles were also removed. Full texts of 210 papers were assessed for eligibility. A total of 33 studies were finally included in this review. Details of the studies, including study design, study objective, health professionals involved, measurements, SNSs evaluated, and conclusions, are summarized in Multimedia Appendix 1. The studies (n=177) that were excluded are shown in Multimedia Appendix 2, along with the reasons for their exclusion.

Characteristics of the Reviewed Studies

Among the 33 included studies, more than half of the reviewed studies (n=19) were published in recent 2 to 3 years (between 2015 and 2017). The studies were conducted in 11 countries, the majority being based in the United Kingdom (n=9), the United States (n=12), and Canada (n=4). Other countries with one study included were Australia, China, France, Israel, Saudi Arabia, Malaysia, Singapore, and Turkey (Multimedia Appendix 1). Participants in the reviewed studies were from diverse health professions (Table 1). On many occasions, more than one health profession was involved in the studies evaluating the use of SNSs in clinical practice. Two studies were conducted in large multidisciplinary communities of practice [18,19]. Physicians, including medical and surgical doctors, were involved in about two-thirds of the studies (n=19). Students and trainees were involved in 7 studies in which the uses of SNSs in education and training were evaluated [20-26].

Assessing the Quality of the Studies

Overall, the quality of studies was satisfactory. Most of the reviewed studies met the criteria in checklists (Multimedia Appendix 3). All studies were exploratory in nature, and the findings were often descriptive. Among the 33 studies, 12 were quantitative [20,21,27-36], 5 qualitative [18,37-40], and 16 used mixed methods [19,22-26,41-50].

No randomized clinical trials (RCTs) were included in this review. No head-to-head comparisons of the relative effectiveness of SNSs could be identified. Researchers often used more than one approach in examining the roles of SNSs and their outcomes. Most studies used surveys (n=25) [19-23,25-36,42-45,47-50]. The questionnaires adopted in the surveys were mostly developed by the researchers. No validated scale was used for surveying the use of SNSs among health professionals. Therefore, conducting a meta-analysis was not possible in this review.

In most of the mixed methods studies, researchers conducted surveys and then analyzed the messages (or communication) in the SNSs. This method is called “content analysis.” Researchers also analyzed the characteristics of SNS users and the context of their communications and SNSs metrics, such as the number of messages, posts, tweets, likes, and followers. Five studies conducted one-on-one interviews [24,37,39,41,46], and one used focus group interviews [44]. Thematic analysis was used in these studies, with key themes being identified from the content of the communications (eg, WhatsApp messages) and user comments.

Uses and Benefits of Social Network Sites for Professional Communication

The 33 included studies involved a range of SNSs. In 11 studies, the authors conducted cross-sectional surveys or interviews to examine participants’ utilization of any types of social media and SNSs in the broad sense, without concentrating on any particular type of SNS. Among the rest of papers, the most reported SNSs are Twitter, Facebook, WhatsApp Messenger, and LinkedIn. Table 2 describes the types of SNSs studied. All the studies investigating the use of Twitter and Facebook were conducted in North America and the United Kingdom, and those studying WhatsApp Messenger were based in the United Kingdom, the Middle East, and Asia (Multimedia Appendix 1). The one evaluating Sina Weibo was based in China.

Predictors of use of SNSs for professional purposes were often examined by researchers. The positive predictors identified include younger age (20-39 years), fewer years of professional experience (0-10), and lower rank, such as residents and nonconsultants [36,44,48,50]. All retrieved studies stated that SNSs enhanced effective communication and information sharing among health professionals. Participants in the reviewed studies appreciated SNSs as user-friendly, free, and fast tools for communication [24,31,38,45]. The utilization and benefits of SNSs for communication among health professionals in (1) frontline clinical practice, (2) professional networks, and (3) education and training are examined in the following paragraphs.

Figure 1. Literature search following Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) guidelines. SNS: social network site.

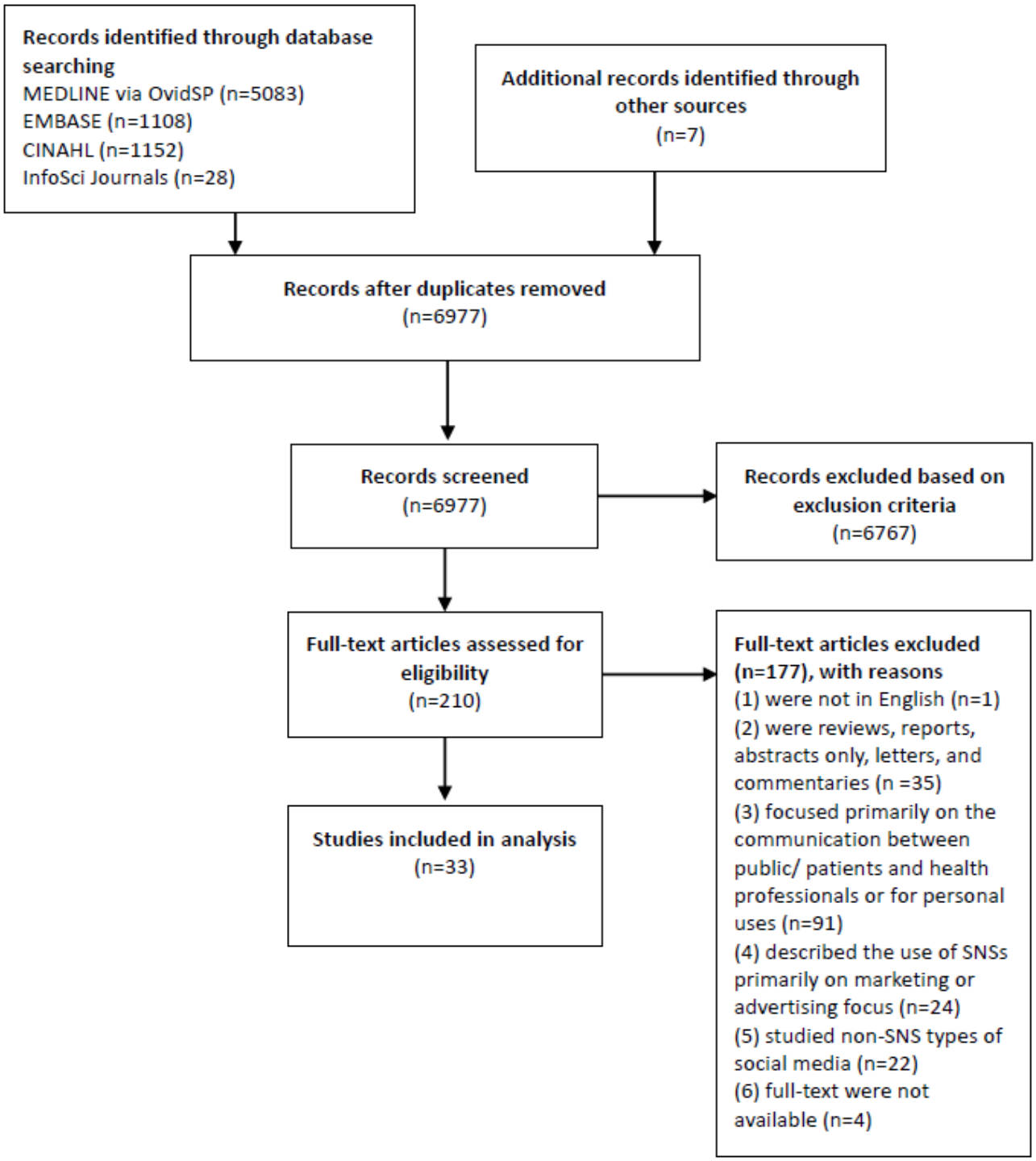


Table 1. Types of health professions included in the reviewed studies.

Health profession	Number of studies[a]
Medical (physicians, specialists, surgeons, and medical students)	20
Pharmacy (pharmacists, pharmacy students, and faculty)	8
Nursing (nurses and student nurses)	4
Multidisciplinary community of practice	2
Forensic occupational therapy	1
Public health	1
Radiology	1

[a]A study could include more than one type of health profession.

Table 2. Types of social network sites.

Tool or app	Number of studies[a]
Twitter	8
Facebook	7
WhatsApp	6
LinkedIn	2
Sina Weibo	1
Yahoo online discussion group	1
Web 2.0 (tools not specified)	1
Any types of social media or network site	11

[a]More than 1 social network site was involved in some studies.

Uses and Benefits of Frontline Clinical Practice

In the delivery and operation of health services, SNSs are used as channels for communication within clinical teams [27,43,46,49], for seeking clinical consultation or making referrals to consultants or specialists [30,38], for disseminating clinical guidelines, and for promoting awareness of the guidelines among practitioners [32]. WhatsApp Messenger was used when instant responses and actions were required within the framework of the same institute [27,38,46,49]. In a multisite family health team involving many members scattered throughout a territory, Facebook was selected as a tool for communication, collaboration, and informal knowledge exchange [43].

The key benefit was that SNSs being the convenient and efficient channels for information sharing had no restriction by locations or office hours. They were effective in creating a complex, longitudinal stream of information and multimedia files [19]. Photographic and diagnostic images, text messages, videos, and voice messages (eg, rhythm sounds in the monitor worn by patients) were easily shared via WhatsApp messages [27,30,38,46,49]. SNSs allowed the sharing of messages with multiple recipients, which shortened the time for processing. In Wani et al's study (2013), participating physicians and consultants commented that WhatsApp Messenger was a fast and effective method for the team to evaluate patients and to complete academic endorsement [49]. It was also claimed that it helped to flatten the hierarchy within a clinical team [46]. Johnston et al (2015) and Wani et al (2013) noted that WhatsApp Messenger continued to be the communication system used within the teams after the completion of studies [46,49].

Uses and Benefits of Professional Networks

SNSs were used to build and strengthen interactions and relationships of a professional nature. They facilitated connections and collaborations among practitioners of the same health profession [30]. SNS users can strategically search for and join groups of their communities or common interests, such as professions and research areas, that enhanced network building among health professionals of diverse backgrounds but with the same interests, connecting them beyond the scope of their usual practices [18,45].

For instance, the formation of impressive networks among Twitter #hcsmca community members not constrained by professional status was revealed in the social network analysis performed by Gruzd and Haythornthwaite (2013) [18]. In Goff et al's study (2016), the Twitter group engaged plentiful professionals interested in infectious diseases and antimicrobial stewardship topics [42]. The LinkedIn group, "Hand Surgery International," demonstrated a remarkable gain in membership, up to 4106 in 4 years. The building of this community of practice took place beyond geographical limitations [45].

The establishment of professional networking and making new contacts was one of the most favorable benefits brought by SNSs [28,37,45]. Professionals can also create a professional online presence, increasing the number of their followers and

having a greater impact on readership and content dissemination [29]. Users have a high level of control over the content that they read, listen to, watch, and follow. Among the interviewees in Benetoli et al's exploratory study (2016), Facebook was preferred over other SNSs for professional purposes because of its popularity, simplicity, and versatility [37].

Uses and Benefits of Education and Training

SNSs were used as novel tools for teaching, learning, and enhancing educational interactions among peers, students, instructors, and preceptors [20-23,25,26,47]. Twitter and Facebook were used in course assignments and projects. They were found to be useful, straightforward educational tools to supplement and enhance students' learning experience [20,21,25]. The utility, feasibility, and acceptability of WhatsApp Messenger in supplementing "problem-based learning" was clearly indicated in the study by Raiman et al (2017) [24]. When used to support teaching and learning, SNSs encouraged interactivity in both peer and academic support [22,23,26]. The applications fostered a positive social atmosphere, generating learning opportunities outside the classroom [24]. They enhanced the construction of students' own learning and the continuation of their engagement in development [20]. Reames et al (2016) concluded that SNSs positively influenced the educational experience and engagement of students [25].

Health professionals can stay abreast of news and information pertaining to their professional interests by following or subscribing to updates in SNSs [21,28,29,33]. For instance, the latest clinical information and real-time surveillance data on an infectious outbreak could be released ahead of peer-reviewed published papers [42]. Facebook and Twitter aided promoting professional development [41,42,50] and also facilitated outreach from a scientific conference, allowing active participation via communication during the conference [19].

Limitations of Social Network Sites for Professional Communication

Some drawbacks come with the utilization of SNSs. How to operate the SNSs smoothly was a challenge to some health professionals [28,29,36]. In Nikiphorou et al's study (2016), 30% of non–social media users justified not using SNSs because of lack of knowledge on how to do so [29]. Patel et al (2017) also pointed to unfamiliarity with the technical aspects of SNSs as one of the obstacles to their utilization.

Hesitations on the use of SNSs included concerns regarding data protection, patient privacy, and liability [28,29,43,44]. In a survey, more than half of the respondents were uncertain regarding the procedures or mechanisms for archiving or backing up data [30]. Although WhatsApp Messenger was successfully integrated into the operations of clinical teams, members were concerned that WhatsApp conversations could be regarded as medical records [49]. Fuoco and Leveridge (2015) raised the controversy of whether medical regulatory bodies should monitor the social media activities of health professionals [31]. Nonetheless, whether there was any institutional policy regarding transfer of personal medical information by SNSs was seldom mentioned in studies.

The border between the professional and personal spheres of SNS use was blurred to many health professionals [39]. Exposure of one's private life was one of the risks of using SNSs that contain detailed personal profile [22,29]. Some health professionals had concerns over the stigma of unprofessionalism and a negative impact on their reputation from the use of SNSs [28,29]. Academic faculty members worried whether being "friended" on Facebook or "followed" on Twitter would blur the boundaries of the instructor-student relationship [33]. On the other side, students said that they felt revision anxiety because their module leaders could read about their personal lives on Facebook [22]. SNSs often provide instant messaging functions. Concerns about the intrusiveness and pushiness of messages, particularly after office hours, were raised by members of clinical teams that used WhatsApp Messengers [24,38].

The implementation of SNSs was not found to be beneficial or effective to participants in all the reviewed studies. Although more than 80% of students agreed that the Sina Weibo improved communication, one-fourth felt that collaborative learning was not effective [26]. Reluctant participation was observed in the use of Twitter designed for enhancing the educational experience of a clerkship. Only 8% of respondents (5 of 62) agreed that Twitter could increase their clerkship engagement [25]. It was proposed that the reluctance was due to the one-way flow of information. In Maisonneuve et al's study (2015), participants checked (or read) SNS content more often than they posted, and the exchanges on SNSs were limited [39]. Gruzd and Haythornthwaite (2013) concluded that leadership and members' participation were crucial for the effectiveness of online networks [18].

Discussion

Principal Findings

The 33 included studies in this review provided evidence that SNSs have been developed as useful platforms for communication among health professionals with significant benefits in the frontline clinical practice, professional networks, and education and training.

Numerous benefits of using SNSs were identified. SNS users in the reviewed studies considered SNSs as user-friendly, easy-to-use, free, and fast tools for communication [24,31,38,45]. In frontline clinical practice, SNSs were efficient in transferring a stream of information and multimedia files instantly to multiple recipients. This highly facilitated the communication among members of the service units or teams [19,27,30,38,46,49]. In building professional networks, SNSs connected professionals beyond the scope and geographical locations of their usual practices [18,42,45]. Users were benefited in making new contacts and expanding their networks [28,37,45]. As tools for education and training, SNSs were useful in generating learning opportunities and enhancing interactions among peers, students, instructors, and preceptors [20-23,25,26,47]. They also promoted update of news and professional development [21,28,29,33,41,42,50].

The merit of SNSs in facilitating interactions, sharing of information, and promoting connections among health professionals is well illustrated in this review. Compared with the findings of the reviews that examined the uses of SNSs between the public and health professionals, this review added value by summarizing the benefits of SNSs in communication among health professionals [9-11,51]. With the increasing use of SNSs, there will be further opportunities to use this efficient tool for professional communication.

In this review, the most reported SNSs were Twitter, Facebook, WhatsApp Messenger, and LinkedIn. Twitter and LinkedIn are robust in expanding a user's connection because users can easily follow their targets without disclosing much private details or requesting authorization [5,6]. Facebook is designed to share one's personal profile with "friends"; hence, it may disclose more personal details. It was best used when building and strengthening a community among a group of known people, such as members of the National Physicians Alliance [41], and large cohorts of students [21]. WhatsApp Messenger was appraised as an efficient and easy-to-use app for communication in clinical teams or for linking up students and instructors [24,27,30,38,46,49]. However, its use was constrained within an established framework or a group of recipients because users' mobile phone numbers must be sought to join a group.

The positive predictors of SNSs uses identified by the included studies were younger age, fewer years of professional experience, and lower rank [36,44,48,50]. Other reviews on the use of SNSs in health communication and education also revealed that young people intend to use SNSs more than the older ones [11,51]. This observation aligns with the current profiles of SNS users such as 59% of active Facebook users are between the ages of 18 and 34 years [52]. It would warrant research exploring how this batch of "SNSs-competent" students would influence health communication when they come into practice in the near future.

In addition to the requirements on technical knowledge [28,29,36], the uncertainties on data protection and liability were also obstacles to the utilization of SNSs [28,29,43,44]. Moreover, the blurred border between the professional and personal spheres [39] and the risk of exposing one's private life imposed further hesitation on using SNSs [22,29].

As the growth of SNSs is expected to rise, health professions should have a better understanding of how to attain secure and appropriate use of these platforms. Formal training should be provided to health professionals for the safe use of SNSs [33]. The American Medical Association recommends that physicians consider separating personal and professional information online, and they preserve professional boundaries when interacting with patients [53]. In a survey involving clerkship directors in the United States, most respondents felt that a faculty member accepting a friend request from a current student was never or rarely appropriate [54]. Yet, guidance on faculty-student or faculty-trainee interactions, particularly when SNSs are used as an educational tool, is often inadequate. Academic faculty could find it confusing to maintain appropriate boundaries in the instructor-student relationship [33].

Those concerns over the possible stigma and the negative impact of reputation on the use of SNSs fall within the context of e-professionalism. It is defined as the attitudes and behaviors that reflect traditional professionalism paradigms but are manifested through digital media [55,56]. E-professionalism is an essential and increasingly important element of professional identity formation [56]. Discussion on this topic helps to preserve the integrity of health professions, establish appropriate boundaries, and protect the privacy of both patients and professionals [57]. Unexpectedly, the relevant discussion was limited in most of the studies in this review. Evolving challenges are expected with the emerging use of SNSs; e-professionalism should be included in the education of health professionals and incorporated in institute policy and staff training.

Although not much mentioned in the included papers, a practical issue that should be given attention is how the SNS companies manage, analyze, repurpose, or even disclose the data and content of communication. According to the terms of service of Facebook, Twitter, WhatsApp Messenger, and LinkedIn, the companies reserve the right to collect, use, preserve, and share users' information if it is deemed reasonably necessary to respond to legal process, government requests, or to enforce the companies' terms and policies, and also under a list of other situations [58-61]. Health service institutes and providers must consider carefully in using SNSs for communicating confidential data to avoid jeopardizing patient privacy.

Gaps in the Literature and Potential Areas for Further Research

In this review, the first key observation was the absence of an RCT among the included studies. All studies were exploratory in nature. The majority used surveys, content analysis, and thematic analysis. This illustrated the early phase of research in the field of professional use of SNSs when researchers were more concerned with describing health professionals' behavior and opinions rather than the effectiveness of SNSs itself. The number of retrieved studies has risen considerably in the last 5 years, and it is expected to see significant growth in the research on SNSs soon. When this area of research advances further, research design will likely progress to interventional study. Some potential study designs are cross-sectional study, longitudinal study, and RCTs. It is worth mentioning that, nowadays, many analytics tools for SNSs are being developed in the market. Researchers can analyze straightforwardly how the content and performances of SNS interventions are affecting the study outcomes.

Every SNS is unique in design, interfaces, uses, and target users. To compare the relative effectiveness of SNSs for communication among health professionals, further research with more robust methodologies such as RCTs would be required. For instance, an RCT was conducted to investigate a physical activity intervention with pedometers delivered via Facebook app [62]. Another RCT was conducted to compare interventions via the WhatsApp Messenger and the Facebook social group in preventing smoking relapse in quitters [63].

With the emerging use of SNSs, evolving challenges in the context of e-professionalism are expected. This topic should be covered in the education of health professionals and incorporated

in institute policy and staff training. Concerning the data policies of the SNSs companies, institutes must consider carefully in using SNSs for sharing confidential data. Research investigating the mechanisms of data protection and the potential risks in sharing information in SNSs should be conducted to identify suitable ways for safe use and maintenance of data.

Geographical locations may affect the generalization of findings in research on SNSs. The availability, acceptability, and popularity of SNSs vary across countries and populations. Twitter, among the top 3 SNSs in the United States, ranked ninth in Hong Kong, with only 10% of market share [64], whereas QZone, the top SNS in China, may not be heard by many Americans [8]. For higher applicability of findings to local practice, research has to be done in the corresponding location and jurisdiction. If published data are inadequate, exploratory study designs such as cross-sectional survey, preferably together with a validation study, should be conducted to explore health professionals' perceptions, the barriers, and usage patterns of SNSs in professional communication. This helps to pave the way for research on more robust methodologies.

An effective and sustainable online network is crucial for the communication via SNSs. As discussed in the reviewed studies, not all the implementations of SNSs were found to be beneficial or effective [25,26,39]. Research could be done to explore strategies for designing and enhancing the usability of SNSs in communication among health professionals.

Limitations of the Review

The absence of RCT coupled with the diverse and heterogeneous designs of the included studies has made conducting a meta-analysis unfeasible. Most studies were surveys and interviews, and their measurements and findings were mostly descriptive and qualitative. In addition, the questionnaires adopted in the surveys were mostly developed by the researchers, where validation might not be done. There were often some questions in common, such as asking respondents to distinguish the use of SNSs for personal or for professional purposes. Yet, the definitions of personal versus professional use of SNSs varied across studies. Without a well-stated explanation of terms, questions were sometimes ambiguous, for example, "How have you used or benefited from social media professionally?" [28].

The definitions of social media and SNSs could be unclear in some earlier literature. This complicates the analysis of the primary studies in this review. Another related limitation was the keyword search. The term "social network site" has not been added to the MeSH list in PubMed. This issue was addressed by undertaking a series of searches using a range of keywords, such as the names of common SNSs. Nevertheless, the searches may not have captured all relevant publications. Research on SNSs is growing so fast that evidence may have been published in electronic media or platforms not indexed through the academic databases. Thus, findings in this review are limited to research published in traditional peer-reviewed journals only.

Conclusions

SNSs provide platforms facilitating efficient communication among health professionals in frontline clinical practice, professional networks, and education and training. Disseminating information, expanding professional connections, and promoting interactions are the benefits observed. Yet, the advantages come with limitations such as requirements on technical knowledge, professionalism issues, and risks of data protection. The evolving use of SNSs necessitates further robust research to explore the full potential and relative effectiveness of SNSs in professional communication.

Acknowledgments

The authors wish to acknowledge the anonymous reviewers for their thoughtful review and guidance. There was no funding for this study.

Authors' Contributions

WSYC contributed to data collection, data analysis, interpretation of the findings, and drafting the paper. AYML contributed to the study design, data analysis, interpretation of the findings, intellectual input, and revision of the paper. Both authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Summary of the reviewed studies.

[PDF File (Adobe PDF File), 72KB - jmir_v20i3e117_app1.pdf]

Multimedia Appendix 2

List of the excluded studies and the reasons for exclusion.

[PDF File (Adobe PDF File), 89KB - jmir_v20i3e117_app2.pdf]

Multimedia Appendix 3

Quality of the reviewed studies.

[PDF File (Adobe PDF File), 391KB - jmir_v20i3e117_app3.pdf]

References

1. Boyd DM, Ellison NB. Social network sites: definition, history, and scholarship. J Comput Mediat Commun 2007 Oct;13(1):210-230 [FREE Full text] [doi: 10.1111/j.1083-6101.2007.00393.x] [Medline: 24843434]
2. Obar JA, Wildman S. Social media definition and the governance challenge: an introduction to the special issue. Telecomm Policy 2015 Oct;39(9):745-750. [doi: 10.1016/j.telpol.2015.07.014]
3. Statista. 2017. Global social media ranking 2017 URL: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/ [accessed 2017-10-30] [WebCite Cache ID 6uaiXTjAF]
4. Facebook. Company Info URL: https://newsroom.fb.com/company-info/ [WebCite Cache ID 6uailwyUQ]
5. Twitter. 2017. URL: https://about.twitter.com/company [accessed 2017-10-30] [WebCite Cache ID 6uaiqoJXd]
6. LinkedIn. 2017. About LinkedIn URL: https://press.linkedin.com/about-linkedin [accessed 2017-10-30] [WebCite Cache ID 6uaiuDDwR]
7. WhatsApp Messenger. Whatsapp. 2017. About WhatsApp URL: http://www.whatsapp.com/about/ [accessed 2017-10-30] [WebCite Cache ID 6uaixyM9Y]
8. World Economic Forum. 2017. The world's most popular social networks, mapped URL: https://www.weforum.org/agenda/2017/03/most-popular-social-networks-mapped/ [accessed 2017-10-30] [WebCite Cache ID 6uajL2X2a]
9. Capurro D, Cole K, Echavarría MI, Joe J, Neogi T, Turner AM. The use of social networking sites for public health practice and research: a systematic review. J Med Internet Res 2014 Mar;16(3):e79 [FREE Full text] [doi: 10.2196/jmir.2679] [Medline: 24642014]
10. 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]
11. Laranjo L, Arguel A, Neves AL, Gallagher AM, Kaplan R, Mortimer N, et al. The influence of social networking sites on health behavior change: a systematic review and meta-analysis. J Am Med Inform Assoc 2015 Jan;22(1):243-256. [doi: 10.1136/amiajnl-2014-002841] [Medline: 25005606]
12. Rolls K, Hansen M, Jackson D, Elliott D. How health care professionals use social media to create virtual communities: an integrative review. J Med Internet Res 2016 Jun 16;18(6):e166 [FREE Full text] [doi: 10.2196/jmir.5312] [Medline: 27328967]
13. Coiera E. Social networks, social media, and social diseases. Br Med J 2013;346:f3007. [Medline: 23697672]
14. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Br Med J 2009 Jul 21;339:b2535. [doi: 10.1136/bmj.b2535]
15. Critical Appraisal Skills Programme. 2017. CASP Tools & Checklists URL: http://www.casp-uk.net/casp-tools-checklists [accessed 2017-10-30] [WebCite Cache ID 6uajMu8CU]
16. National Collaborating Centre for Methods & Tools. 2011 Dec 16. Critical appraisal tools to make sense of evidence URL: http://www.nccmt.ca/resources/search/87 [accessed 2017-10-30] [WebCite Cache ID 6uajOi05C]
17. National Heart, Lung, and Blood Institute. 2017. Quality Assessment Tool for Observational Cohort Cross-Sectional Studies URL: https://www.nhlbi.nih.gov/health-pro/guidelines/in-develop/cardiovascular-risk-reduction/tools/cohort [accessed 2017-10-30] [WebCite Cache ID 6uajQQGuh]
18. 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]
19. 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]
20. Desselle SP. The use of Twitter to facilitate engagement and reflection in a constructionist learning environment. Curr Pharm Teach Learn 2017 Mar;9(2):185-194. [doi: 10.1016/j.cptl.2016.11.016]
21. Kostka-Rokosz MD, Camiel LD, McCloskey WW. Pharmacy students' perception of the impact of a Facebook-delivered health news service-Two-year analysis. Curr Pharm Teach Learn 2014 Jul;6(4):471-477. [doi: 10.1016/j.cptl.2014.04.007]
22. Mawdsley A, Schafheutle EI. Using Facebook to support learning and exam preparation in a final-year undergraduate pharmacy clinical therapeutics module. Curr Pharm Teach Learn 2015 Nov;7(6):869-875. [doi: 10.1016/j.cptl.2015.08.010]
23. Morley DA. Supporting student nurses in practice with additional online communication tools. Nurse Educ Pract 2014 Jan;14(1):69-75. [doi: 10.1016/j.nepr.2013.06.005] [Medline: 23871299]
24. Raiman L, Antbring R, Mahmood A. WhatsApp messenger as a tool to supplement medical education for medical students on clinical attachment. BMC Med Educ 2017 Jan 06;17(1):7 [FREE Full text] [doi: 10.1186/s12909-017-0855-x] [Medline: 28061777]

25. Reames BN, Sheetz KH, Englesbe MJ, Waits SA. Evaluating the use of Twitter to enhance the educational experience of a medical school surgery clerkship. J Surg Educ 2016 Jan;73(1):73-78. [doi: 10.1016/j.jsurg.2015.08.005] [Medline: 26417905]
26. Wang T, Wang F, Shi L. The use of microblog-based case studies in a pharmacotherapy introduction class in China. BMC Med Educ 2013 Sep 08;13:120 [FREE Full text] [doi: 10.1186/1472-6920-13-120] [Medline: 24010945]
27. Ganasegeran K, Renganathan P, Rashid A, Al-Dubai SA. The m-Health revolution: exploring perceived benefits of WhatsApp use in clinical practice. Int J Med Inform 2017 Jan;97:145-151. [doi: 10.1016/j.ijmedinf.2016.10.013] [Medline: 27919374]
28. Patel SS, Hawkins CM, Rawson JV, Hoang JK. Professional social networking in radiology: who is there and what are they doing? Acad Radiol 2017 May;24(5):574-579. [doi: 10.1016/j.acra.2016.09.026] [Medline: 28153576]
29. Nikiphorou E, Studenic P, Ammitzbøll CG, Canavan M, Jani M, Ospelt C, EMEUNET. Social media use among young rheumatologists and basic scientists: results of an international survey by the Emerging EULAR Network (EMEUNET). Ann Rheum Dis 2017 Dec;76(4):712-715. [doi: 10.1136/annrheumdis-2016-209718] [Medline: 27797750]
30. Siegal G, Dagan E, Wolf M, Duvdevani S, Alon EE. Medical information exchange: pattern of global mobile messenger usage among otolaryngologists. Otolaryngol Head Neck Surg 2016 Nov;155(5):753-757. [doi: 10.1177/0194599816656178] [Medline: 27352892]
31. Fuoco M, Leveridge MJ. Early adopters or laggards? Attitudes toward and use of social media among urologists. BJU Int 2015 Mar;115(3):491-497. [doi: 10.1111/bju.12855] [Medline: 24981237]
32. Narayanaswami P, Gronseth G, Dubinsky R, Penfold-Murray R, Cox J, Bever C, et al. The impact of social media on dissemination and implementation of clinical practice guidelines: a longitudinal observational study. J Med Internet Res 2015;17(8):e193 [FREE Full text] [doi: 10.2196/jmir.4414] [Medline: 26272267]
33. Cain J, Scott DR, Tiemeier AM, Akers P, Metzger AH. Social media use by pharmacy faculty: student friending, e-professionalism, and professional use. Curr Pharm Teach Learn 2013 Jan;5(1):2-8. [doi: 10.1016/j.cptl.2012.09.002]
34. Deen SR, Withers A, Hellerstein DJ. Mental health practitioners' use and attitudes regarding the Internet and social media. J Psychiatr Pract 2013 Nov;19(6):454-463. [doi: 10.1097/01.pra.0000438184.74359.88] [Medline: 24241499]
35. Keller B, Labrique A, Jain KM, Pekosz A, Levine O. Mind the gap: social media engagement by public health researchers. J Med Internet Res 2014;16(1):e8 [FREE Full text] [doi: 10.2196/jmir.2982] [Medline: 24425670]
36. Stevens RJ, Hamilton NM, O'Donoghue JM, Davies MP. The use of the Internet and social software by plastic surgeons. Eur J Plast Surg 2012 Mar 15;35(10):747-755. [doi: 10.1007/s00238-011-0681-z]
37. Benetoli A, Chen TF, Schaefer M, Chaar BB, Aslani P. Professional use of social media by pharmacists: a qualitative study. J Med Internet Res 2016 Sep 23;18(9):e258 [FREE Full text] [doi: 10.2196/jmir.5702] [Medline: 27663570]
38. Gulacti U, Lok U, Hatipoglu S, Polat H. An analysis of WhatsApp usage for communication between consulting and emergency physicians. J Med Syst 2016 Jun;40(6):130. [doi: 10.1007/s10916-016-0483-8] [Medline: 27083574]
39. Maisonneuve H, Chambe J, Lorenzo M, Pelaccia T. How do general practice residents use social networking sites in asynchronous distance learning? BMC Med Educ 2015 Sep 21;15:154 [FREE Full text] [doi: 10.1186/s12909-015-0435-x] [Medline: 26391989]
40. Dieleman C, Duncan EA. Investigating the purpose of an online discussion group for health professionals: a case example from forensic occupational therapy. BMC Health Serv Res 2013 Jul 03;13:253 [FREE Full text] [doi: 10.1186/1472-6963-13-253] [Medline: 23822895]
41. Flynn S, Hebert P, Korenstein D, Ryan M, Jordan WB, Keyhani S. Leveraging social media to promote evidence-based continuing medical education. PLoS One 2017 Jan;12(1):e0168962 [FREE Full text] [doi: 10.1371/journal.pone.0168962] [Medline: 28060854]
42. Goff DA, Jones C, Toney B, Nwomeh BC, Bauer K, Ellison EC. Use of Twitter to educate and engage surgeons in infectious diseases and antimicrobial stewardship. Infect Dis Clin Pract 2016;24(6):324-327. [doi: 10.1097/IPC.0000000000000440]
43. Lofters AK, Slater MB, Nicholas Angl E, Leung FH. Facebook as a tool for communication, collaboration, and informal knowledge exchange among members of a multisite family health team. J Multidiscip Healthc 2016 Jan;9:29-34 [FREE Full text] [doi: 10.2147/JMDH.S94676] [Medline: 26869796]
44. Barry AR, Pearson GJ. Professional use of social media by pharmacists. Can J Hosp Pharm 2015;68(1):22-27. [Medline: 25762816]
45. Dong C, Cheema M, Samarasekera D, Rajaratnam V. Using LinkedIn for continuing community of practice among hand surgeons worldwide. J Contin Educ Health Prof 2015;35(3):185-191. [doi: 10.1002/chp.21300] [Medline: 26378424]
46. Johnston MJ, King D, Arora S, Behar N, Athanasiou T, Sevdalis N, et al. Smartphones let surgeons know WhatsApp: an analysis of communication in emergency surgical teams. Am J Surg 2015 Jan;209(1):45-51. [doi: 10.1016/j.amjsurg.2014.08.030] [Medline: 25454952]
47. Lipp A, Davis RE, Peter R, Davies JS. The use of social media among health care professionals within an online postgraduate diabetes diploma course. Pract Diabetes 2014 Jan 28;31(1):14-17a. [doi: 10.1002/pdi.1821]
48. Loeb S, Bayne CE, Frey C, Davies BJ, Averch TD, Woo HH, American Urological Association Social Media Work Group. Use of social media in urology: data from the American Urological Association (AUA). BJU Int 2014 Jun;113(6):993-998 [FREE Full text] [doi: 10.1111/bju.12586] [Medline: 24274744]

49. Wani SA, Rabah SM, Alfadil S, Dewanjee N, Najmi Y. Efficacy of communication amongst staff members at plastic and reconstructive surgery section using smartphone and mobile WhatsApp. Indian J Plast Surg 2013 Sep;46(3):502-505 [FREE Full text] [doi: 10.4103/0970-0358.121990] [Medline: 24459338]
50. Wang AT, Sandhu NP, Wittich CM, Mandrekar JN, Beckman TJ. Using social media to improve continuing medical education: a survey of course participants. Mayo Clin Proc 2012 Dec;87(12):1162-1170 [FREE Full text] [doi: 10.1016/j.mayocp.2012.07.024] [Medline: 23141117]
51. Benetoli A, Chen TF, Aslani P. The use of social media in pharmacy practice and education. Res Social Adm Pharm 2015 Feb;11(1):1-46. [doi: 10.1016/j.sapharm.2014.04.002] [Medline: 24814268]
52. Statista. 2017. Distribution of Facebook users worldwide as of January 2018, by age and gender URL: https://www.statista.com/statistics/376128/facebook-global-user-age-distribution/ [accessed 2017-10-30] [WebCite Cache ID 6uajW7bav]
53. American Medical Association. AMA. Professionalism in the use of social media URL: https://www.ama-assn.org/sites/default/files/media-browser/public/about-ama/councils/Council%20Reports/council-on-ethics-and-judicial-affairs/i10-ceja-professionalism-use-social-media.pdf [accessed 2018-03-16] [WebCite Cache ID 6xxn7COHC]
54. Chretien KC, Farnan JM, Greysen SR, Kind T. To friend or not to friend? Social networking and faculty perceptions of online professionalism. Acad Med 2011 Dec;86(12):1545-1550. [doi: 10.1097/ACM.0b013e3182356128] [Medline: 22030752]
55. Cain J, Romanelli F. E-professionalism: a new paradigm for a digital age. Curr Pharm Teach Learn 2009 Dec;1(2):66-70. [doi: 10.1016/j.cptl.2009.10.001]
56. Kaczmarczyk JM, Chuang A, Dugoff L, Abbott JF, Cullimore AJ, Dalrymple J, et al. e-Professionalism: a new frontier in medical education. Teach Learn Med 2013;25(2):165-170. [doi: 10.1080/10401334.2013.770741] [Medline: 23530680]
57. Spector ND, Matz PS, Levine LJ, Gargiulo KA, McDonald MB, McGregor RS. e-Professionalism: challenges in the age of information. J Pediatr 2010 Mar;156(3):345-346. [doi: 10.1016/j.jpeds.2009.12.047] [Medline: 20176180]
58. LinkedIn Corporation. Linkedin. 2017. User Agreement URL: https://www.linkedin.com/legal/user-agreement [accessed 2018-02-27] [WebCite Cache ID 6xXP3fk4g]
59. Twitter. 2017. Twitter Terms of Service: Privacy URL: https://twitter.com/en/tos#usPrivacy [accessed 2017-10-30] [WebCite Cache ID 6uajahlUT]
60. WhatsApp Messenger. Whatsapp. 2017. Terms of Service URL: http://www.whatsapp.com/legal/ [accessed 2017-10-30] [WebCite Cache ID 6uajcrEdg]
61. Facebook. 2017. Data Policy URL: https://www.facebook.com/about/privacy/ [accessed 2017-10-30] [WebCite Cache ID 6uajkuoBp]
62. Maher C, Ferguson M, Vandelanotte C, Plotnikoff R, De Bourdeaudhuij I, Thomas S, et al. A web-based, social networking physical activity intervention for insufficiently active adults delivered via Facebook app: randomized controlled trial. J Med Internet Res 2015;17(7):e174 [FREE Full text] [doi: 10.2196/jmir.4086] [Medline: 26169067]
63. Cheung YT, Chan CH, Lai CK, Chan WF, Wang MP, Li HC, et al. Using WhatsApp and Facebook online social groups for smoking relapse prevention for recent quitters: a pilot pragmatic cluster randomized controlled trial. J Med Internet Res 2015;17(10):e238 [FREE Full text] [doi: 10.2196/jmir.4829] [Medline: 26494159]
64. Go-globe. 2015. Social Media Usage in Hong Kong URL: https://www.go-globe.hk/blog/social-media-hong-kong/ [accessed 2017-11-06] [WebCite Cache ID 6ulL1oaNi]

Abbreviations

CASP: Critical Appraisal Skills Programme
MeSH: medical subject heading
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analyses
RCT: randomized clinical trial
SNS: social network site

Edited by CL Parra-Calderón; submitted 20.07.17; peer-reviewed by L Laranjo, K Ganasegeran, R Alkoudmani, A Gilavand, IN Gomez, A Cyr; comments to author 11.09.17; revised version received 06.11.17; accepted 18.01.18; published 28.03.18.

Please cite as:
Chan WSY, Leung AYM
Use of Social Network Sites for Communication Among Health Professionals: Systematic Review
J Med Internet Res 2018;20(3):e117
URL: http://www.jmir.org/2018/3/e117/
doi:10.2196/jmir.8382
PMID:29592845

Original Paper

The Optimization of an eHealth Solution (Thought Spot) with Transition-Aged Youth in Postsecondary Settings: Participatory Design Research

Nicole VanHeerwaarden[1], MHI; Genevieve Ferguson[1], MEd; Alexxa Abi-Jaoude[1], MPH; Andrew Johnson[1], BA; Elisa Hollenberg[1], MSW; Gloria Chaim[2,3], MSW; Kristin Cleverley[2,4], RN, PhD; Gunther Eysenbach[5,6], MPH, MD; Joanna Henderson[2,3], PhD; Andrea Levinson[3,7], MSc, MD; Janine Robb[8], MSc; Sarah Sharpe[9], PhD; Aristotle Voineskos[3,10], MD, PhD; David Wiljer[1,3,11], PhD

[1]Education, Centre for Addiction and Mental Health, Toronto, ON, Canada
[2]McCain Centre for Child, Youth & Family Mental Health, Centre for Addiction and Mental Health, Toronto, ON, Canada
[3]Department of Psychiatry, University of Toronto, Toronto, ON, Canada
[4]Faculty of Nursing, University of Toronto, Toronto, ON, Canada
[5]Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada
[6]Centre for Global eHealth Innovation, University Health Network, Toronto, ON, Canada
[7]Early Intervention Clinic, Centre for Addiction and Mental Health, Toronto, ON, Canada
[8]Health and Wellness Centre, University of Toronto, Toronto, ON, Canada
[9]QoC Health, Toronto, ON, Canada
[10]Slaight Family Centre for Youth in Transition, Centre for Addiction and Mental Health, Toronto, ON, Canada
[11]Education, Technology & Innovation, University Health Network, Toronto, ON, Canada

Corresponding Author:
David Wiljer, PhD
Education, Technology & Innovation
University Health Network
190 Elizabeth Street
R Fraser Elliott Building RFE 3S-411
Toronto, ON, M5G 2C4
Canada
Phone: 1 416 340 6322
Email: david.wiljer@uhn.ca

Abstract

Background: Seventy percent of lifetime cases of mental illness emerge before the age of 24 years, but many youth are unable to access the support and services they require in a timely and appropriate way. With most youth using the internet, electronic health (eHealth) interventions are promising tools for reaching this population. Through participatory design research (PDR) engagement methods, Thought Spot, a Web- and mobile-based platform, was redeveloped to facilitate access to mental health services by transition-aged youth (aged 16-29 years) in postsecondary settings.

Objective: The aim of this study was to describe the process of engaging with postsecondary students through the PDR approaches, with the ultimate goal of optimizing the Thought Spot platform.

Methods: Consistent with the PDR approaches, five student-led workshops, attended by 41 individuals, were facilitated to obtain feedback regarding the platform's usability and functionality and its potential value in a postsecondary setting. Various creative engagement activities were delivered to gather experiences and opinions, including semistructured focus groups, questionnaires, personas, journey mapping, and a world café. Innovative technological features and refinements were also brainstormed during the workshops.

Results: By using PDR methods of engagement, participants knew that their ideas and recommendations would be applied. There was also an overall sense of respect and care integrated into each group, which facilitated an exchange of ideas and suggestions.

Conclusions: The process of engaging with students to redesign the Thought Spot platform through PDR has been effective. Findings from these workshops will significantly inform new technological features within the app to enable positive help-seeking behaviors among students. These behaviors will be further explored in the second phase that involves a randomized controlled trial.

(J Med Internet Res 2018;20(3):e79) doi:10.2196/jmir.8102

KEYWORDS

students; transition-aged youth; mental health; substance use; eHealth; mobile apps; participatory action research; help-seeking

Introduction

Background

The transition between childhood and adulthood can be difficult and many transition-aged youth will seek information about mental health and wellness (for the purposes of this study, we define transition-aged youth as those aged 16-29 years). Within Canada, mental health is a significant concern for young adults, with rates of mood disorders (8%) and substance use disorders (12%) higher among 15- to 24-year-olds than any other age group [1]. A Canadian survey of youths' Web-based resource preferences showed that 52% of respondents aged 16 years to 25 years had previously sought information about mental illness symptoms, 47% had sought information about treatment, and 24% had sought Web-based questionnaires or assessment tests related to mental health and substance use [2]. An Australian survey reported similar findings, with one-third of 18- to 25-year-olds reporting primarily depending on the internet for information about mental health or substance use problems [3]. Given the increased use of Web-based resources as sources of mental health information, electronic health (eHealth) platforms are effective and promising options for delivering reliable information and improving access to mental health and wellness services for transition-aged youth. This study focuses on transition-aged youth in postsecondary settings. The terms *postsecondary students* or *students* will be used throughout this paper to describe our target population.

What Is Thought Spot?

Thought Spot is a crowdsourced digital platform (mobile- and Web-based) that aims to better enable transition-aged youth in postsecondary settings to seek and access mental health and wellness services. It was developed by the Centre for Addiction and Mental Health and the University of Toronto (UT), with partners' Ryerson University (RU), the Ontario College of Art and Design, and ConnexOntario. The project was funded by the Ontario Ministry of Training, Colleges and Universities [4]. Thought Spot is a student-led project that prioritizes inclusion through steering committees, working groups, and focus groups. Postsecondary students were involved in initial decisions about the project name, logo, product design, and project management. Through cross-organizational collaboration between postsecondary students and project partners, Thought Spot became a platform that invites students to share their knowledge about services, discover wellness options in their area, and read reviews of services. Using an interactive and crowdsourced map, users are able to geo-locate mental health and wellness spots. All spots are categorized by the type of services offered, and users can apply filters to personalize their search. Evaluation data collected during the first phase of Thought Spot showed that students felt a sense of ownership over the product because of their contributions, gained knowledge in the areas of mental health and wellness, and developed new skills throughout their involvement that were transferable to their education and future careers [4]. These data helped to inform the second phase of the Thought Spot project, which will be discussed in this paper.

This study includes two new stages: (1) optimizing the Thought Spot platform by engaging with students; and (2) measuring the impact of Thought Spot on help-seeking behaviors of students in postsecondary settings through a randomized controlled trial [4]. The objective of this paper is to describe the participatory design research (PDR) methods used during the optimization phase of the project and summarize the results. This study explored how PDR should be employed in designing and optimizing mHealth interventions for student mental health, as well as a discussion of the utility of the various data-gathering techniques.

Methods

Participatory Action and Design Research

The optimization of Thought Spot was conducted using PDR methodologies while following some of the principles of participatory action research (PAR) [4]. PAR is "a social, collaborative learning process" [5] that involves an iterative process of engaging end users in reflection to provide a deeper understanding of their needs and experiences [5-7]. Participants involved in such projects are empowered to work alongside researchers as equal contributors [8]. The goal of PAR is to include all stakeholders throughout the entire process [9] and to work toward and implement solutions that target clearly defined problems [10]. In this case, participants were asked to work on a specific problem that had already been established through PDR methods and, therefore, there were limitations to the extent to which PAR was applied. Similar to PAR, PDR involves the target audience in codesigning the technologies that audience will use [4]. In eHealth research, a number of techniques can be used to implement PDR, including workshops, ethnography, prototyping, and user-design activities [9]. PDR is most effective when the design of the intervention is driven by the values of the stakeholders [9]. PDR is based on actively engaging participants to take an equal role in developing and designing a product or service around their own experiences [11]. Although the principles of PAR and PDR align, each methodology relies on slightly different techniques. For example, PDR focuses on the design of a product or technology [4,11], whereas PAR focuses more on the process of research [9].

Various techniques and tools are used to generate understanding of the experiences and needs of end users. Methods in this eHealth project included two-part discussions (a large group discussion followed by a small group activity), semistructured interviews, questionnaires, personas, journey maps, and world cafés. The existing literature on PAR and PDR informed the structure of our activities with participants.

Recruitment

A total of 41 participants attended 5 workshops. Participants were current students or recent graduates from the UT, RU, and George Brown College (GBC). To encourage students with lived experience to participate, explicit wording on the recruitment poster was used: "students with lived experience of mental health and substance use are encouraged to participate." Workshops took place between July and September 2016.

Numerous methods were used to recruit participants for our engagement workshops. The workshops were promoted through preexisting Thought Spot social media accounts: Twitter, Facebook, and Instagram. Recruitment flyers were posted on departmental boards at UT, RU, and GBC. Academic departments and student organizations were identified as potential recruitment sites, including departments of psychology and social work, and health and wellness centers. The existing connections within the Thought Spot student advisory group and the research team were also used to help recruit participants. Participants received a small honorarium and public transit tokens for attending each workshop, and food was served at all workshops. Participants also received a list of mental health and addictions resources they could access, if needed. All participants signed an informed consent form that provided an overview of the study objectives, risks, benefits, confidentiality, and contact information. This study obtained research ethics board approval from the Centre for Addiction and Mental Health, UT, RU, and GBC.

Data Collection

Usefulness, Satisfaction, and Ease of Use Questionnaire

To assess participants' opinions on the platform's usefulness and satisfaction, the Usefulness, Satisfaction, and Ease of Use (USE) questionnaire was distributed to all participants [12] following each workshop. The USE questionnaire is a standardized scale consisting of 30 quantitative questions divided into 4 main sections: Usefulness, Ease of Use, Ease of Learning, and Satisfaction. Questions are asked using a 7-point scale from strongly disagree to strongly agree. The data collected informed the rebuild of the Thought Spot platform.

Codesign Workshops

A total of 5 workshops were planned based on the codesign activity methods identified through preliminary research. Each activity used different elements of PDR techniques to elicit information through a range of group formats (Table 1). Sociodemographic information and a postworkshop evaluation survey were collected at each workshop. All workshops were audio-recorded and flip chart notes collected. Workshops were facilitated by a research coordinator with 7 years of experience in facilitating focus groups with vulnerable populations, a research analyst with over 5 years of experience in facilitating focus groups through alternative methods of engagement, and 3 practicum students interested in cocreation. Facilitators used a semistructured question guide tailored for each activity to guide the discussion. Sample images from flip chart notes taken during the workshops can be found in Multimedia Appendix 1.

Various facilitation techniques were used in the workshops. A semistructured approach to facilitating group discussions established flexibility while maintaining an overall sense of direction throughout the conversation between participants [13]. All workshops were organized into 2 distinct parts. Two-part discussions allowed for an initial general conversation about the identified topic or question, followed by a more targeted conversation [14]. This approach gives participants the opportunity to discuss general topics outside the context of the product being researched. For example, in one of our two-part discussions, questions in the first portion of the discussion focused on the general experience of accessing mental health and wellness services as a student, without any focus on Thought Spot. Participants were then divided into smaller groups to further explore and discuss barriers to seeking help related to mental health and wellness in a more intimate setting. Mazzone et al reported that when engaging with youth, small groups "allow for greater focus on each task" while fostering creativity. Dividing participants into smaller groups during discussions helped ensure that most participants were able to contribute [15,16].

Table 1. Workshop descriptions.

Workshop #	# of participants	Structure	Purpose
1	6	Semistructured, two-part discussion	Explore the usage of eHealth apps and gain insights into the Thought Spot user experience from experienced users.
2	8	Semistructured, two-part discussion with small breakout groups	Explore the use of eHealth apps and gain insights into the Thought Spot user experience from new users.
3	8	Semistructured, two-part discussion using personas	Determine whether Thought Spot meets the health needs of its user personas.
4	6	User journey mapping followed by semistructured discussion	Explore the experiences of new users through journey mapping.
5	13	Focus group followed by a world café	Gather information on what health needs Thought Spot addresses, what features to include in its redesign, and what would keep users coming back.

Our team used personas, world café, and journey mapping as methods for gathering information from participants. Personas are "realistic descriptions of a type of client or user" that help to establish an understanding of the needs and perspectives of those for whom a product is being designed [17]. The process of walking through the experiences of users helps to guide and focus improvements for specific products or services [17]. Personas provide an appropriate amount of structure that allows participants to communicate ideas in the context of a larger topic in a tangible way [18,19].

In our study, several personas were developed to help capture diverse backgrounds and to outline different scenarios one might encounter when looking to access mental health or wellness services in the Greater Toronto Area (GTA). An example of one of our personas can be found in Multimedia Appendix 2. The use of personas maintained a level of confidentiality by inviting participants to discuss how Thought Spot could meet their needs without needing to disclose their own personal experiences.

Journey mapping was another technique used for collecting information. It is commonly used to evaluate the user experience through accessing and interacting with a service or product over time [17]. Suggestions for improving health interventions or accessing programs have been uncovered through journey mapping as participants highlight specific points of contact within the health care system and the emotions they experience while navigating it [11,17]. A world café exercise was used during the final Thought Spot workshop, which focused on confirming our findings from the previous four workshops and eliciting diverse perspectives. World cafés involve small groups cycling through a series of questions at different stations and building on the answers of the previous groups [20,21]. This method attempts to obtain diverse perspectives, rather than to achieve consensus, to better understand the overall experience of participants [20]. By dividing into small groups, participants had an additional opportunity to express their opinions about Thought Spot and share their overall experience of help seeking. Finally, participants who were uncomfortable discussing sensitive topics in a group setting could give written feedback through questionnaires [15].

Workshop Evaluation Feedback

Feedback surveys were completed by participants following each workshop. The surveys collected information on how participants heard about the workshop and what they liked and disliked about it, as well as asking participants whether they had any additional questions or comments about the project in general. This feedback was used to guide adjustments to subsequent workshops.

Data Analysis

Usefulness, Satisfaction, and Ease of Use Questionnaire Analysis

IBM SPSS Statistics 24 was used to analyze the USE questionnaire data (N=27). Values for low (1-2), medium (3-5), and high (6-7) satisfaction in the USE questionnaire were calculated by taking the sum of responses for each question.

Inductive Content Analysis

The data collected during the workshops were analyzed using content analysis, a method often used "for making replicable and valid inferences from data to their context, with the purpose of providing knowledge, new insights, a representation of facts, and a practical guide to action" [22]. An inductive approach to content analysis was chosen, as no preconceived set of categories or framework was used to design the study [22]. This method is advantageous because the topics discussed came directly from participants [23].

The method of inductive content analysis involves 3 main phases: preparation, organizing, and reporting [24]. The preparation phase involves selecting the object of study for the content analysis, known as a unit of analysis [25]. Although the unit of analysis can come in many forms, whole interviews or observational protocols are most suitable [25]. The organizing phase involves the open coding of transcripts, generating a list of topics, and grouping similar topics together to form categories [24]. A process of abstraction then occurs, whereby a general description of the research topic is created from the groups of topics [24]. During the final, reporting phase, a model or conceptual map is generated to present the results [24].

Audio recordings of the workshops were sent to a professional transcriptionist. The transcripts were then anonymized and sent to all participants for review. The varied facilitation techniques (breakout groups, small/large group discussions) resulted in 16 transcripts produced from the 5 workshops. The units of analysis were transcripts of whole workshops. Transcripts of activities within each workshop were combined to create a single transcript for each workshop. Combining all transcripts from each workshop resulted in 5 discrete transcripts and ensured that data collected from each workshop were equally prioritized. To identify key discussion topics, 2 researchers independently coded a sample of 3 transcripts. A coding meeting was held where both researchers compared interpretations of the transcripts. The topics were compared and combined into categories of content topics in a coding matrix (Multimedia Appendix 3). Following the creation of the coding matrix, the 3 full transcripts were coded by each researcher to test its validity. Once the coding matrix was verified, content analysis of all transcripts was performed using QSR International NVivo 10 for Windows qualitative analysis software by one of the researchers.

Results

Demographics

In total, 41 students participated in the workshops: 29 females and 12 males. Most participants were aged between 19 and 24 years (Table 2). Most participants were full-time university or college students (n=39). Of all participants, 2 participants were attending school part-time. More than half of participants indicated that they had some experience with mental health or substance use concerns (Table 2).

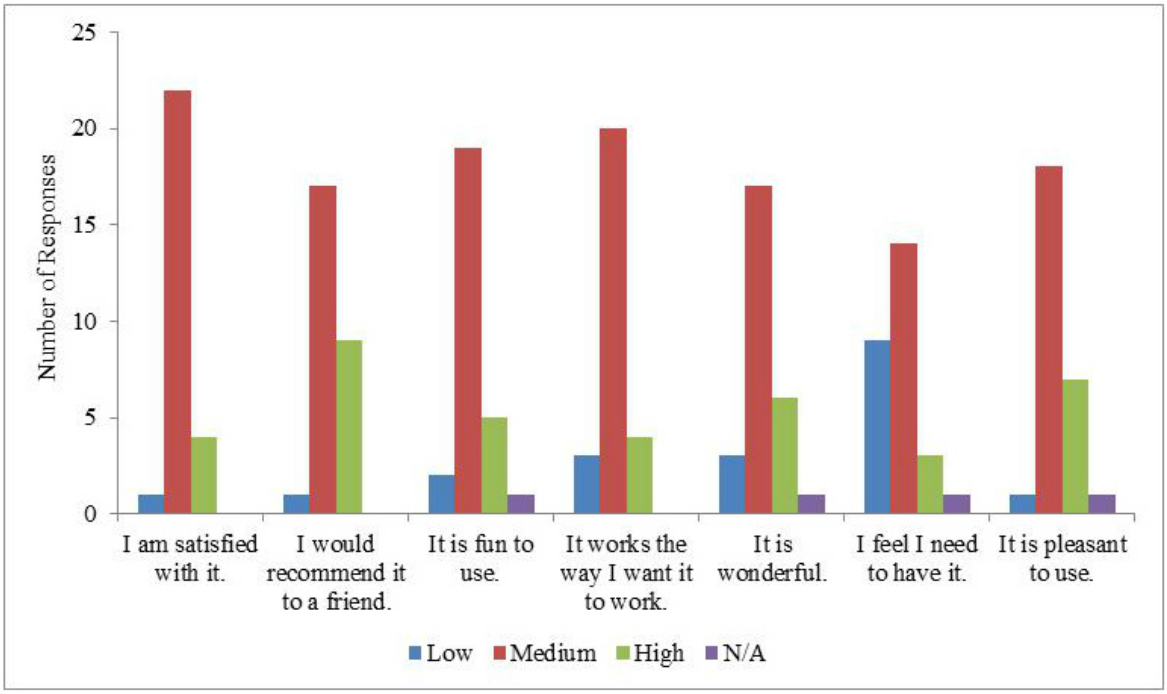
Usefulness, Satisfaction, and Ease of Use Questionnaire

The majority of respondents indicated a medium level of satisfaction with original version of Thought Spot (Figure 1).

Table 2. Participant characteristics (n=41).

Participant characteristic	Percentage (%)
Age (year)	
19-21	24
22-24	63
25-27	7
28-29	3
Other	3
Experience with mental health and/or addiction issues	
Yes	54
No	41
Don't know	5

Figure 1. Participant responses to the satisfaction questions from Usefulness, Satisfaction, and Ease of Use Questionnaire. N/A: not applicable.



Qualitative Codesign Workshop Data

The results from the workshops were reviewed by the core research and design team to inform the redesign of Thought Spot. As each workshop focused on different topics and the facilitators used various data collection methods, the content that was coded varied. A high-level concept map was created based on the coding of topics across all transcripts (Figure 2). In Figure 2, larger circles represent topics that were discussed more often during the workshops.

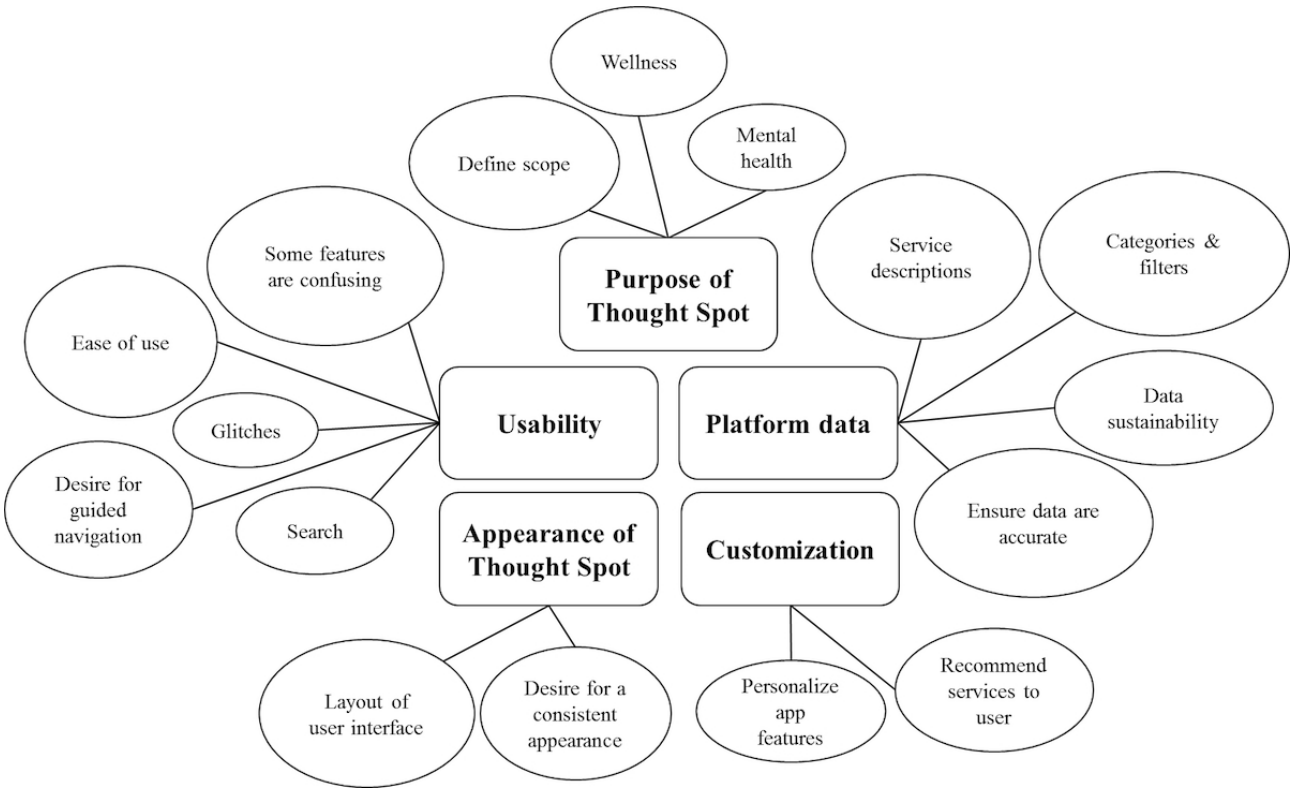
Purpose of Thought Spot

Participants often disagreed on the purpose of Thought Spot. Some thought that recreational programs, social clubs, and tips for maintaining mental well-being should be included in the platform. For others, however, limiting the focus to mental health services seemed fitting. Further discussions explored this difference in opinion, and workshop facilitators explained the purpose of Thought Spot as a wellness app that recognized the broad determinants of health, mental health, and wellness.

When discussing different types of wellness activities and services that could be included in Thought Spot, participants also mentioned preventative approaches such as connecting users to student social groups or building a function for tracking emotions, moods, and thoughts. One participant stated:

> *The best way to treat mental health is through mental wellness. Prevention is the best policy. That also speaks about some of the transitions; if you're starting something new [like starting university or college], you might want to be able to have access to community services like yoga classes or support groups.*

Figure 2. High-level conceptual map of topics discussed during workshops.



Participants who supported framing Thought Spot as a wellness app believed that this would increase its overall accessibility to a broader range of people. Although there was a general consensus on incorporating a wellness approach in the platform, participants encouraged the research team to carefully consider the inclusion criteria for these services.

Usability

Usability was a concern for some participants. Certain features of Thought Spot were described as confusing and difficult to use. For example, adding a spot was particularly difficult for participants, as one participant explained:

> *I found it very difficult to try to add anything...I tried everything 3 times and it would freeze or shut down.*

Participants mentioned that using the categories or filters to find services was challenging due to the confusing categorical structure or a lengthy list of filters. For example, they found the category "Health and Social Services" to be too broad because it included youth drop-in centers as well as community produce markets and community parks. Participants suggested including descriptions of categories and tutorials that walk users through each feature in the app. Some participants felt that, in general, the app was easy to use due to its similarity to other geo-location apps such as Google Maps and Yelp. One participant stated:

> *I think the way Google Maps does it is useful for me with the TTC [public transit] or walking routes, things like that.*

Participants requested that the navigation feature of Thought Spot be expanded to include in-app directions to spots.

Platform Data

Discussions about data varied with each workshop group. Topics included the categories and filters, crowdsourcing, description of services, missing information, forming partnerships, ratings, sustainability, and data verification. As Thought Spot is a crowdsourced app, participants identified active moderating as a method of maintaining a clean dataset. One participant suggested:

> *Anyone can add something, so maybe there could be a way to confirm that these are valid. With a checkmark or something that shows that this has been verified by someone on the back-end.*

Participants recommended that a member from the research team with experience and interest in moderation and data cleansing be responsible for this process. This moderator would fill in missing information (eg, address, hours) and populate description fields. Participants requested that descriptions of spots include details such as cost, appointment or walk-in, accessibility, hours of operation, parking, and available languages.

Appearance of Thought Spot

Discussions about the appearance of Thought Spot often focused on the layout. Some participants felt that the interface was too cluttered and overwhelming. For example, the resources page was described as "really dense" with "a lot of text and it's just black and white, so it's not pretty." Most feedback was about having too much information displayed on each screen.

Generally, participants liked the overall consistency with the color schemes and design layout, but some participants highlighted inconsistencies in how the app is displayed on

various devices (iPads, iPhones, Android). Participants agreed that a consistent use of color, shapes, and layout of features was ideal. Comments about the app's color scheme were positive.

Participants responded well to the "friendly and approachable color scheme." In the words of one participant:

> *You have a lot of greens and softer lilacs and blues going on. It feels like a health and wellness app.*

Customization

Overall, participants were in favor of being able to customize Thought Spot to each user's specific needs and preferences. The participants proposed the ability to save a list of favorite spots and the option to personalize different features. For example, some participants requested the ability to modify general settings such as sounds or number of push notifications they received. Participants also discussed the possibility of developing Thought Spot as a smart app, which would provide recommendations to users based on their unique needs, interests, and search history. Participants also recommended that suggested spots could be based on the user's mood. One participant suggested:

> *Maybe just have an option that allows you to put in how you feel that day, or things that concern you that you might want to talk about. Then it will take in those things and suggest certain services or certain people that you can reach out to.*

Workshop Evaluation Feedback

After each workshop, participants were sent a short feedback survey containing 5 questions. These evaluations indicated that participants enjoyed the collaborative, interactive environment and felt safe sharing their thoughts and opinions. Our team received conflicting feedback about the duration of the workshops. For some participants, ensuring that adequate time was allocated to each portion of the workshops was a key concern, whereas others thought the workshops could have been shorter. Participants mentioned that time spent filling out surveys (sociodemographic and USE questionnaire) could have been better spent with group discussion, and that surveys could be filled out before the workshops. Of the 41 evaluation feedback surveys distributed, 25 were completed.

Discussion

Target Population Composition

Comparing our participants with Ontario postsecondary students in general, we see some similarities and differences. The majority of our participants were female (71%, 29/41), and 95% (39/41) were full-time postsecondary students compared with 55% and 80% for all Ontario students in the 2015 and 2016 school year, respectively [26]. Moreover, 87% (36/41) of our sample were between the ages of 19 and 24 years, whereas Statistics Canada reports that 46% of Ontario postsecondary students are between the ages of 20 and 24 years [27]. Finally, 54% (22/41) of participants indicated they had lived experience of mental health and/or substance use which is higher than the reported Canadian average for this population [2]. Given that this is a qualitative study that relied on self-selection of a small sample size, we did not anticipate recruiting a fully representative sample of our target population.

Optimization of Thought Spot

Our team focused on eliciting qualitative and quantitative feedback on how to improve the first version of Thought Spot. There was an overall interest among participants in helping to develop an mHealth intervention that streamlines access to mental health and wellness services for their peers. Results from both the qualitative data analysis and USE questionnaire show a moderate level of satisfaction with the current Thought Spot platform. When assessing its usability, participants discussed the features they found confusing to use, ambiguity surrounding the categories and filters, and the desire to ensure that the information about services is accurate and up to date.

At times the feedback from students conflicted and therefore presented challenges for the project team to make design decisions. For example, during discussions about whether wellness-type services should be included in the platform, some students supported the idea, but others disagreed, wanting the platform to focus solely on mental health services. Conflicting opinions were taken into account during the redesign process. Cost, timelines, and capacity to implement some of the suggestions also had to be thoughtfully weighed by the project team.

To help guide the design process, a design working group was established that included research team members, technological partners, and student representatives. This group discussed and prioritized the needs and wants identified by the students who participated in the workshops. Design decisions were also brought back to the Thought Spot Student Group, our advisory group, for feedback and confirmation.

Use of Participatory Design Research Methods

The success of this project to date supports the move toward PDR in the area of mHealth interventions targeted toward transition-aged youth [28]. Fundamental to PDR is the need to involve target users in all aspects of the research and to empower them to have a sense of ownership over the product. Although it may be difficult to include participants as equal members of the research team, efforts should be made to ensure that their views are valued and embedded into the product design whenever possible. The strength of these research methods lies in ensuring that an open collaboration between researchers and participants exists. Using PDR during this optimization phase of the project created an environment in which our participants were encouraged to contribute their experiences and ideas related to health, mental health, and wellness to ultimately improve the Thought Spot platform.

Motivating students to engage in PDR requires fostering a sense of understanding of the approach and allowing participants to engage in a way that maintains confidentiality and safety [19]. To address potential power imbalances between participants and researchers, the facilitators ensured that good communication and respect between these 2 groups were established at the beginning of each workshop [29]. Cocreating a series of workshop guidelines with participants was essential to ensure that the workshop environment was open,

collaborative, and safe. Workshop evaluations indicated that participants enjoyed the collaborative, interactive environment, and that they felt welcomed to share their thoughts and experiences. The facilitators made a concerted effort to foster an environment where differing opinions and experiences were valued and encouraged by actively listening to each participant by facilitating discussion so all participants could contribute thoughts and ideas.

Various PDR methods were used to elicit feedback on participant's opinions of Thought Spot from the perspective of postsecondary student's experiences of mental health and wellness. The techniques found to be most useful during the Thought Spot workshops were small group discussions, persona exercises, and journey mapping. The use of personas was previously used by Nicholas and colleagues to help research participants discuss youth-specific challenges [19]. We found similar benefits when using this technique in that imagining a best outcome for the persona increased a sense of ownership and empathy in participants [19]. The use of journey maps during the workshops gave participants the opportunity to develop and express a narrative about their experiences with the platform. In addition to facilitating discussion of user experiences, the journey maps prompted discussion of the appearance and purpose of Thought Spot. The journey mapping exercise proved to be very effective in identifying areas for improvement. This could be due to the open-ended format of journey mapping, where participants were able to provide feedback that was not limited by questions asked by facilitators.

These creative techniques made the workshops accessible to participants, helped them relate to the topic of mental health, and encouraged them to contribute to solutions [19]. Collaborating with students in PDR has significant benefits, including establishing common ground and understanding the needs and motivations of the target population [19]. Each PDR technique used during the workshops assisted researchers in collecting helpful feedback for optimizing the Thought Spot platform. Understanding the unique strengths of each method to answer specific questions or collect different types of feedback is critical to the success of PDR. Thoughtful consideration of what techniques to use in a PDR project can help to ensure that the desired feedback is collected.

Limitations

Participants represented only 3 postsecondary campuses in the GTA, all of which are located in downtown Toronto. The experiences of these students' mental health help seeking may vary from those studying and living in other parts of the city where services are less accessible. Although efforts were made to recruit more males, the majority of our participants were female, potentially skewing our data. Another challenge involves the methodology used in our study. Using various methods to gather information in each workshop meant that slightly different data were collected. In addition, a relatively small number of participants (N=25) completed the USE questionnaire, and descriptive data analysis was performed by the research team.

Conclusions

Students encounter barriers to seeking help, such as confusion when navigating the health system and fears of being labeled. Services made available through a crowdsourced platform may facilitate and enhance the help-seeking process. Moderate satisfaction with the current Thought Spot platform can be improved by addressing concerns with usability, content accuracy, and customization. PDR methods are useful tools when engaging students in research related to eHealth. PDR is most effective when the design is driven by the values of the stakeholders [9]. The values expressed by students have guided Thought Spot's platform optimization and redesign. Engaging with students through in-person workshops and activities was very effective for this project. The redesign of Thought Spot was guided by feedback received through these PDR workshops. Next steps include testing the effectiveness of the platform through a randomized controlled trial and continuing to enhance the overall project operations based on feedback received from student participants.

Acknowledgments

The authors would like thank the following funders: the Ontario Ministry of Training, Colleges and Universities funded the initial development of Thought Spot through the Mental Health Innovation Fund; and the Canadian Institutes of Health Research, through the eHealth Innovations Partnership Program, funded this study. The authors also thank the Centre for Addiction and Mental Health project team that developed the initial Thought Spot prototype: Miriam Verburg, Michelle Hamilton-Page, Annie Hart, Doug Giles, and Lawrie Korec. The authors also thank the universities, colleges, and partners that have contributed to this project: University of Toronto, Ryerson University, George Brown College, OCAD University, University Health Network's Centre for Global eHealth Innovation, ConnexOntario, Kids Help Phone, and mindyourmind. The authors thank QoC Health and the Slaight Centre for Youth in Transition for their outstanding contributions and support. The authors would like to thank Hema Zbogar for her editing expertise on this manuscript, and to Mandee Kaur and Chelsea Stunden for organizing and cofacilitating the summer workshops. Finally, the authors express deep gratitude to all those who have generously contributed, and continue to contribute, their energy, enthusiasm, ideas, and passion in cocreating and coproducing this work to improve the experiences and lives of postsecondary students.

Conflicts of Interest

SS is a cofounder and shareholder in QoC Health. GE is the editor-in-chief and publisher of the *Journal of Medical Internet Research* but was not involved in the peer-review process or decision-making for this paper.

Multimedia Appendix 1

Example images from workshops.

[PDF File (Adobe PDF File), 423KB - jmir_v20i3e79_app1.pdf]

Multimedia Appendix 2

Thought Spot persona example.

[PDF File (Adobe PDF File), 380KB - jmir_v20i3e79_app2.pdf]

Multimedia Appendix 3

Coding matrix of topics discussed during Thought Spot workshops.

[PDF File (Adobe PDF File), 216KB - jmir_v20i3e79_app3.pdf]

References

1. Pearson C, Janz T, Ali J. Ottawa, ON: Statistics Canada; 2013. Mental health and substance use disorders in Canada. Health at a Glance series URL: http://www.statcan.gc.ca/pub/82-624-x/2013001/article/11855-eng.pdf [accessed 2017-05-26] [WebCite Cache ID 6qkTZyAl7]
2. Wetterlin FM, Mar MY, Neilson EK, Werker GR, Krausz M. eMental health experiences and expectations: a survey of youths' Web-based resource preferences in Canada. J Med Internet Res 2014;16(12):e293 [FREE Full text] [doi: 10.2196/jmir.3526] [Medline: 25519847]
3. Burns JM, Davenport TA, Durkin LA, Luscombe GM, Hickie IB. The internet as a setting for mental health service utilisation by young people. Med J Aust 2010 Jun 7;192(11 Suppl):S22-S26. [Medline: 20528703]
4. Wiljer D, Abi-Jaoude A, Johnson A, Ferguson G, Sanches M, Levinson A, et al. Enhancing self-efficacy for help-seeking among transition-aged youth in postsecondary settings with mental health and/or substance use concerns, using crowd-sourced online and mobile technologies: the thought spot protocol. JMIR Res Protoc 2016 Nov 4;5(4):e201 [FREE Full text] [doi: 10.2196/resprot.6446] [Medline: 27815232]
5. Bennett S, Whitehead M, Eames S, Fleming J, Low S, Caldwell E. Building capacity for knowledge translation in occupational therapy: learning through participatory action research. BMC Med Educ 2016 Oct 1;16(1):257 [FREE Full text] [doi: 10.1186/s12909-016-0771-5] [Medline: 27716230]
6. Goeman D, Michael J, King J, Luu H, Emmanuel C, Koch S. Partnering with consumers to develop and evaluate a Vietnamese Dementia Talking-Book to support low health literacy: a qualitative study incorporating codesign and participatory action research. BMJ Open 2016 Sep 26;6(9):e011451 [FREE Full text] [doi: 10.1136/bmjopen-2016-011451] [Medline: 27670516]
7. Orlowski SK, Lawn S, Venning A, Winsall M, Jones GM, Wyld K, et al. Participatory research as one piece of the puzzle: a systematic review of consumer involvement in design of technology-based youth mental health and well-being interventions. JMIR Hum Factors 2015;2(2):e12 [FREE Full text] [doi: 10.2196/humanfactors.4361] [Medline: 27025279]
8. MacDonald C. Understanding participatory action research: a qualitative research methodology option. Can J Action Res 2012;13(2):34-50.
9. Kang M, Choo P, Watters C. Design for experiencing: participatory design approach with multidisciplinary perspectives. Procedia Soc Behav Sci 2015 Feb;174:830-833. [doi: 10.1016/j.sbspro.2015.01.676]
10. Barrington DJ, Sridharan S, Saunders SG, Souter RT, Bartram J, Shields KF, et al. Improving community health through marketing exchanges: a participatory action research study on water, sanitation, and hygiene in three Melanesian countries. Soc Sci Med 2016 Dec;171:84-93. [doi: 10.1016/j.socscimed.2016.11.003] [Medline: 27836512]
11. Boyd H, McKernon S, Mullin B, Old A. Improving healthcare through the use of co-design. N Z Med J 2012 Jun 29;125(1357):76-87. [Medline: 22854362]
12. Lund AM. Researchgate. 2001. Measuring Usability with the USE Questionnaire URL: https://www.researchgate.net/publication/230786746_Measuring_usability_with_the_USE_questionnaire [accessed 2017-05-26] [WebCite Cache ID 6qkVLshVD]
13. Cross R, Warwick-Booth L. Using storyboards in participatory research. Nurse Res 2016 Jan;23(3):8-12. [doi: 10.7748/nr.23.3.8.s3] [Medline: 26793981]
14. Giesbrecht EM, Miller WC, Mitchell IM, Woodgate RL. Development of a wheelchair skills home program for older adults using a participatory action design approach. Biomed Res Int 2014;2014:172434 [FREE Full text] [doi: 10.1155/2014/172434] [Medline: 25276768]
15. Winterling J, Wiklander M, Obol CM, Lampic C, Eriksson LE, Pelters B, et al. Development of a self-help web-based intervention targeting young cancer patients with sexual problems and fertility distress in collaboration with patient research partners. JMIR Res Protoc 2016 Apr 12;5(2):e60 [FREE Full text] [doi: 10.2196/resprot.5499] [Medline: 27073057]

16. Mazzone E, Read J, Beale R. Design with and for disaffected teenagers. New York, NY, USA: ACM; 2008 Presented at: Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges; 2008; Lund, Sweden p. 290-297. [doi: 10.1145/1463160.1463192]
17. Boyd H, McKernon S, Old A. Auckland, New Zealand: Waitemata District Health Board; 2010. Health service co-design: Working with patients to improve healthcare services URL: http://www.healthcodesign.org.nz/tools/TOOLKIT_WEB_DOWNLOAD.pdf [accessed 2017-05-26] [WebCite Cache ID 6qkU5GM61]
18. Hagen P, Rowland N. 2011. Enabling codesign URL: http://johnnyholland.org/2011/11/enabling-codesign/ [accessed 2017-05-26] [WebCite Cache ID 6qkUL2Wat]
19. Nicholas M, Hagen P, Rahilly K, Swainston N. Using participatory design methods to engage the uninterested. New York: ACM; 2012 Presented at: Proceedings of the 12th Participatory Design Conference: Exploratory Papers, Workshop Descriptions, Industry Cases - Volume 2; August 12-16, 2012; Roskilde, Denmark p. 121-124. [doi: 10.1145/2348144.2348183]
20. MacFarlane A, Galvin R, O'Sullivan M, McInerney C, Meagher E, Burke D, et al. Participatory methods for research prioritization in primary care: an analysis of the World Café approach in Ireland and the USA. Fam Pract 2017 Jun 1;34(3):278-284. [doi: 10.1093/fampra/cmw104] [Medline: 27677298]
21. Stöckigt B, Teut M, Witt CM. AM use and suggestions for medical care of senior citizens: a qualitative study using the World Café method. Evid Based Complement Alternat Med 2013;2013:951245 [FREE Full text] [doi: 10.1155/2013/951245] [Medline: 24023586]
22. Krippendorff K. Content Analysis: An Introduction to It's Methodology. Beverly Hills, CA: Sage; 2004.
23. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
24. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
25. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
26. Statistics Canada. Postsecondary enrolments by institution type, registration status, province and sex (both sexes) URL: http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/educ71a-eng.htm [accessed 2018-02-20] [WebCite Cache ID 6xNhEgxzE]
27. Statistics Canada. Postsecondary enrolments, by program type, credential type, age groups, registration status and sex URL: http://www5.statcan.gc.ca/cansim/a47 [accessed 2018-02-20] [WebCite Cache ID 6xNhHH4GF]
28. Delman J. Participatory action research and young adults with psychiatric disabilities. Psychiatr Rehabil J 2012;35(3):231-234. [doi: 10.2975/35.3.2012.231.234] [Medline: 22246121]
29. Wadley G, Lederman R, Gleeson J, Alvarez-Jimenez M. Participatory design of an online therapy for youth mental health. 2013 Presented at: Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration; November 25-29, 2013; Adelaide, Australia p. 517-526. [doi: 10.1145/2541016.2541030]

Abbreviations

eHealth: electronic health
GBC: George Brown College
GTA: Greater Toronto Area
PDR: participatory design research
PAR: participatory action research
RU: Ryerson University
USE: Usefulness, Satisfaction, and Ease of Use
UT: University of Toronto

Edited by J Torous; submitted 26.05.17; peer-reviewed by M Wiklander, JP Allem; comments to author 29.08.17; revised version received 15.01.18; accepted 05.02.18; published 06.03.18.

Please cite as:
VanHeerwaarden N, Ferguson G, Abi-Jaoude A, Johnson A, Hollenberg E, Chaim G, Cleverley K, Eysenbach G, Henderson J, Levinson A, Robb J, Sharpe S, Voineskos A, Wiljer D
The Optimization of an eHealth Solution (Thought Spot) with Transition-Aged Youth in Postsecondary Settings: Participatory Design Research
J Med Internet Res 2018;20(3):e79
URL: http://www.jmir.org/2018/3/e79/
doi:10.2196/jmir.8102
PMID:29510970

Original Paper

Internet-Based Cognitive Behavioral Therapy for Symptoms of Depression and Anxiety Among Patients With a Recent Myocardial Infarction: The U-CARE Heart Randomized Controlled Trial

Fredrika Norlund[1], PhD (Psych); Emma Wallin[2], MSc (Psych); Erik Martin Gustaf Olsson[1], PhD (Psych); John Wallert[1], MSc (Psych); Gunilla Burell[3], PhD; Louise von Essen[1], PhD (Psych); Claes Held[1,4,5], MD, PhD

[1]Clinical Psychology in Healthcare, Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden

[2]Department of Psychology, Uppsala University, Uppsala, Sweden

[3]Family Medicine and Preventive Medicine, Department of Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden

[4]Cardiology, Department of Medical Sciences, Uppsala University, Uppsala, Sweden

[5]Uppsala Clinical Research Center, Uppsala University, Uppsala, Sweden

Corresponding Author:
Fredrika Norlund, PhD (Psych)
Clinical Psychology in Healthcare
Department of Women's and Children's Health
Uppsala University
Akademiska sjukhuset
Uppsala, 75185
Sweden
Phone: 46 737441304
Email: fredrika.norlund@kbh.uu.se

Abstract

Background: Symptoms of depression and anxiety are common after a myocardial infarction (MI). Internet-based cognitive behavioral therapy (iCBT) has shown good results in other patient groups.

Objective: The aim of this study was to evaluate the effectiveness of an iCBT treatment to reduce self-reported symptoms of depression and anxiety among patients with a recent MI.

Methods: In total, 3928 patients were screened for eligibility in 25 Swedish hospitals. Of these, 239 patients (33.5%, 80/239 women, mean age 60 years) with a recent MI and symptoms of depression or anxiety were randomly allocated to a therapist-guided, 14-week iCBT treatment (n=117), or treatment as usual (TAU; n=122). The iCBT treatment was designed for post-MI patients. The primary outcome was the total score of the Hospital Anxiety and Depression Scale (HADS) 14 weeks post baseline, assessed over the internet. Treatment effect was evaluated according to the intention-to-treat principle, with multiple imputations. For the main analysis, a pooled treatment effect was estimated, controlling for age, sex, and baseline HADS.

Results: There was a reduction in HADS scores over time in the total study sample (mean delta=−5.1, P<.001) but no difference between the study groups at follow-up (beta=−0.47, 95% CI −1.95 to 1.00, P=.53). Treatment adherence was low. A total of 46.2% (54/117) of the iCBT group did not complete the introductory module.

Conclusions: iCBT treatment for an MI population did not result in lower levels of symptoms of depression or anxiety compared with TAU. Low treatment adherence might have influenced the result.

Trial Registration: ClinicalTrials.gov NCT01504191; https://clinicaltrials.gov/ct2/show/NCT01504191 (Archived at Webcite at http://www.webcitation.org/6xWWSEQ22)

(J Med Internet Res 2018;20(3):e88) doi:10.2196/jmir.9710

KEYWORDS

eHealth; treatment adherence and compliance; patient acceptance of health care; patient selection; cardiac rehabilitation

Introduction

Background

Symptoms of depression and anxiety are common after an acute myocardial infarction (MI). Approximately 8% to 30% of patients with a recent MI report depressive symptoms [1], and 13% to 60% of patients report anxiety symptoms [2], with anxiety often co-occurring with symptoms of depression [3]. Post-MI symptoms of depression, anxiety or both are associated with an increased risk of adverse cardiac outcomes [2,4] and reduced quality of life [5].

Several pharmacological treatment trials, with and without psychological support, have been found to reduce symptoms of depression and anxiety among patients with acute coronary syndrome [6,7]. Purely psychological treatment studies have also been effective in reducing symptoms of depression and anxiety in patients with coronary heart disease [8]. Effective treatments have been characterized by adopting techniques used in cognitive behavioral therapy (CBT) [9]. To improve access to effective support, increased engagement in eHealth solutions within the cardiac community has been called upon [10], with internet-based CBT (iCBT) representing an eHealth solution that may improve access to acceptable, effective, and cost-effective psychological treatment [11]. iCBT has been found to reduce symptoms of depression and anxiety among adults with common mental health difficulties [12]. In addition, evidence suggests that guided iCBT may improve disease-related functioning and reduce psychological distress in patients with chronic somatic conditions [13]. Furthermore, preliminary evidence suggests that iCBT may reduce symptoms of depression and anxiety for adults with high cardiovascular risk [14]. However, there is limited knowledge regarding the effectiveness and acceptability of iCBT for symptoms of depression and anxiety among MI patients recruited in a clinical setting.

Objectives

The aim of this randomized controlled trial (RCT) was to evaluate the effectiveness of therapist-guided iCBT versus usual care in patients with a recent MI and comorbid symptoms of depression and anxiety.

Methods

Study Design

The U-CARE Heart study is an RCT comparing therapist-guided iCBT with treatment as usual (TAU). A study protocol, including an internal pilot study, has previously been published [15]. Patients (n=239) were recruited from 25 cardiac clinics in Sweden from September 2013 to December 2016. Outcome measurements were collected at baseline (6-10 weeks post-MI) and at post-treatment follow-up (14 weeks post baseline).

The study protocol was approved by the regional ethics committee in Uppsala (2011/217) and registered at ClinicalTrials.gov on January 5, 2012 (NCT01504191). Three protocol design modifications were made during the ongoing trial. First, the inclusion criteria threshold was lowered from ≥10 to >7 on either of the 2 Hospital Anxiety and Depression Scale (HADS) [16] subscale scores (March 5, 2014), to increase the recruitment rate (after having recruited only 7 patients). Second, minor changes were made to the introduction module after completion of the internal pilot trial including the first 20 patients [15]. Third, a mobile device version of the treatment was launched after 63 patients had been randomized to iCBT, representing 53.8% (63/117) of the total allocated to this trial arm (February 29, 2016).

Patients

Inclusion criteria were as follows: (1) <75 years of age, (2) recent MI <3 months, and (3) score >7 on one or both of the 2 HADS subscales. Exclusion criteria were as follows: (1) scheduled for coronary artery bypass surgery, (2) unable to use computer or internet or email or mobile phone, (3) unable to read Swedish, (4) expected to live for <1 year, (5) anticipated to show poor compliance (eg, substance abuse or not showing up to the cardiac nurse visit), (6) self-reported severe depression or suicidal ideation (Montgomery-Asberg Depression Rating Scale-Self Rated [MADRS-S] total score >34 or MADRS-S item 9>3) [17], and (7) participating in another behavioral intervention trial. Patients in both study arms had access to TAU.

Procedure

Patients were identified and screened for eligibility during a routine visit to a cardiac nurse at 1-8 weeks following their MI. Nurses provided brief trial information and logged all consecutive patients matching the inclusion criteria. U-CARE research staff at the coordinating center (Uppsala) called eligible patients to provide further study information. Written information and an informed consent form were sent to patients via postal service. Patients providing informed consent subsequently received an email with a username and password to access a secure internet-based portal to complete the Web-based baseline assessments. Patients reporting symptoms of depression or anxiety >7 on 1 or both of the 2 HADS subscales were randomized to iCBT or TAU. Patients were randomly assigned (stratified by the clinical recruiting center) with a 1:1 allocation, using a computer-generated code. Randomization occurred automatically in the internet-based portal, with patients receiving an email to inform them of condition assignment.

Patients indicating severe depression or suicidal ideation were contacted via phone and referred to appropriate care and excluded from the trial. Patients who did not complete the Web-based baseline or follow-up assessment were reminded by SMS text messages (short message service, SMS), with research staff blind to group allocation telephoning patients who did not complete the assessment within 1 week of receiving the SMS reminder. Paper-and-pencil assessment forms were sent to patients on request or if they were not reached by telephone.

Table 1. Description of the internet-based cognitive behavioral treatment.

Modules	Psychoeducation	Examples of homework assignments
Introduction	The CBT[a] model	Define personal problems and goals
	Common emotional reactions post-MI[b]	
Managing worry	Worry awareness	Exposure for worry with response prevention
	Rational for worry exposure	
Fear and avoidance	Basic principles for fear and exposure	Graded exposure in situations related to cardiac or other fears
	Rational for graded exposures	
Behavioral activation	Vicious circles in depression	Self-monitoring of mood and daily activities
	Rational for behavioral activation	Plan daily activities
Problem solving	Basic problem-solving skills	Apply problem-solving skills
Communication skills	Basic communication skills and relationship-strengthening skills	Apply communication and relationship-strengthening skills
Applied relaxation training	Applied relaxation training protocol	Practice according to relaxation training protocol
Managing negative thoughts	Cognitive restructuring	Self-monitor thoughts and apply cognitive restructuring skills
Coping with insomnia	Sleep hygiene, stimulus control, and sleep restriction	Self-monitor sleep and apply sleep restriction
Values in life	Personal values and quality of life	Formulate personal values and create an action plan according to them
Relapse prevention	Relapse prevention of depression and anxiety	Identify personal preventive strategies

[a]CBT: cognitive behavioral therapy.

[b]MI: myocardial infarction.

Interventions

Internet-Based Cognitive Behavioral Therapy

The treatment consisted of a 14-week, therapist-guided, tailored CBT intervention delivered via a secure internet-based portal (U-CARE-portal). See Multimedia Appendix 1 for a sitemap and Multimedia Appendix 2 for a screenshot of the portal. The treatment was developed by licensed psychologists, in consultation with patients with a history of depression and anxiety post-MI. The treatment included 10 modules with different themes, adapted to MI patients (Table 1). The introduction module was compulsory, and thereafter, patients were able to choose which modules to work with, as informed by previous research suggesting tailored iCBT provides patients with more control while maintaining treatment quality [18]. Each module contained 2 to 4 treatment steps, with each step including a PDF with text-based psychoeducation, and 1 to 2 homework assignments. Patients were recommended to work with 1 step per week during the treatment period. Homework assignments consisted of self-monitoring, skills training, and engagement in exercises based on CBT techniques (Table 1). Modules were considered complete when all homework assignments within a module were sent to the therapist for feedback. In addition, the iCBT treatment included a library with supplementary material and video clips of interviews conducted with post-MI patients concerning coping with common psychological reactions post-MI. Patients also had access to a discussion board where they could communicate with other patients randomized to the treatment arm.

Therapist Support in Internet-Based Cognitive Behavioral Therapy

Each patient was assigned 1 of the 3 available therapists, who were all licensed psychologists specialized in CBT. Each therapist provided feedback on homework assignments via the portal. The purpose of feedback was to express empathy, encourage work with the treatment, and reinforce treatment activity, all of which has been found to correlate with adherence and outcome [19]. Patients were able to contact their therapist at any time, with therapist responses provided within 48 hours. Patients who were inactive for more than 1 week were contacted by their therapist via telephone, with SMS reminders sent if they were unable to be reached via telephone. Motivational interviewing techniques were used during telephone calls to resolve any identified barriers regarding treatment inactivity. Occasionally, telephone calls included explanations regarding treatment module content; however, calls were not therapeutic and focused on working directly with the material. Telephone call duration ranged between 5 and 30 min. Furthermore, technical support provided by research staff (blinded to allocation) was available via telephone and email.

Treatment as Usual

Patients were treated by the local health care system according to international guidelines regardless of treatment allocation. TAU usually includes secondary preventive interventions (eg, information about risk factors and lifestyle changes), cardiac rehabilitation activities (eg, physical exercise), and psychosocial support (eg, counseling if available). Psychotropic medication was not restricted by study participation.

Assessments

Patient Characteristics

Sociodemographic data were obtained from baseline assessments. Medical and risk factor data were obtained from the SWEDEHEART (Swedish Web-system for Enhancement and Development of Evidence-based care in Heart disease Evaluated According to Recommended Therapies, a Swedish nation-wide quality register) databases RIKS-HIA (Register of Information and Knowledge about Swedish Heart Intensive Care Admissions) and SEPHIA (Secondary Prevention after Heart Intensive Care Admission), covering over 90% of all MIs in Sweden [20].

Primary Outcome

HADS-total score (HADS-T) was the primary outcome measure of self-reported symptoms of depression and anxiety, consisting of 14 items divided equally on 2 subscales: anxiety (HADS-A) and depression (HADS-D). Each item is rated on a 4-point Likert scale, resulting in a total score of 42. Higher scores indicate more severe symptoms, with scores above 7 on either subscale indicating mild symptoms [16]. HADS is a reliable and valid measurement of symptom severity and can detect cases of depression and anxiety in different populations [21]. Several studies support the validity of Web-based administration of HADS [22].

Secondary Outcomes

MADRS-S was used to screen for severe depression and suicidal ideation before inclusion and as a secondary outcome measure of self-reported depression [17]. The scale consists of 9 items, with each item rated on a 7-point Likert scale, with a total score of 54. Higher scores indicate a higher level of depressive symptoms. The MADRS-S has adequate psychometric properties administered via both paper-and-pencil assessment and the internet [23].

The Behavioral Activation for Depression Scale-Short Form (BADS-SF) was used as a secondary outcome measure of self-reported symptoms of depression [24]. The scale has 9 items and 2 subscales: avoidance and activation. Each item is rated on a 7-point Likert scale, with a total score of 54. Higher scores indicate less symptoms of depression. The BADS-SF has shown good reliability and validity, predictive validity, and ability to detect clinically relevant changes [24].

Cardiac anxiety was assessed by the Cardiac Anxiety Questionnaire (CAQ) [25]. The scale consists of 18 items and 3 subscales: fear, avoidance, and focus on cardiac-related stimuli and sensations. Each item is rated on a 4-point Likert scale resulting in a total score of 72, with higher scores indicating a higher level of heart-focused anxiety. CAQ has shown reliability and validity among cardiac patients [25].

Adherence was defined as the proportion of treated patients completing the prescribed amount of content within the treatment period [26]. More data on usage and user experience of the intervention were collected, and a detailed analysis of these is presented elsewhere [27].

Statistical Analysis

A statistical analysis plan prepared in line with the CONSORT (Consolidated Standards of Reporting Trials) 2010 statement was completed before the trial database was locked, and treatment allocation was disclosed. No interim analysis was performed. The study had enough patients (n>126) to detect a medium effect size (Cohen's d=0.5) with the power of 80 at alpha level .05.

Descriptive statistics are presented as mean (SD) or count (%) by treatment group, unless otherwise specified.

The main analysis was conducted according to the intention-to-treat (ITT) principle for all outcomes. Multiple linear modeling was used to analyze the treatment effect on outcomes. Treatment allocation was entered as an independent variable, and HADS-T at follow-up was entered as a dependent variable. To achieve increased precision, age, gender, and baseline HADS-T were entered as covariates. In case of a nonsignificant treatment effect from the main analysis, 2 exploratory analyses with the HADS subscales (HADS-A and HADS-D) as separate outcomes were conducted. For the HADS-A analysis, only patients scoring >7 on the HADS-A subscale at baseline were included, with the corresponding selection applied to the HADS-D analysis. Thus, it was possible for patients to be included in both analyses if they score >7 on both subscales at baseline.

ITT analyses were preceded by multiple imputation via chained equations and predictive mean matching [28]. This was done because (1) there were 11.7% (28/239) with missing values in the main outcome, (2) we could not expect values missing completely at random, and (3) preplanned analyses included multiple outcomes. The imputation model included main effects and the following prespecified interactions: age*treatment and sex*treatment. Moreover, 100 imputed datasets were created. The linear model was thereafter fit to each of these datasets, and resulting effect estimates were pooled using Rubin rules [29]. Sensitivity analyses of HADS-T were conducted on observed data. Supplementary analyses of HADS-T were performed based on per protocol (PP) data from all patients who had completed at least one homework assignment. Secondary outcomes were analyzed using ITT only. We report effect estimates as pooled adjusted point estimates (beta) with 95% CI. Paired t tests were performed for all outcomes (baseline vs follow-up) to assess change over time. The relationship between number of completed homework assignments and changes in HADS-T over time was calculated with Spearman rank-order correlation. Statistical significance was set to 5% (2-tailed).

Analyses were performed in R version 3.4.0 (R Foundation for Statistical Computing, Vienna, Austria) [30] using packages base, foreign, ggplot2, mice, miceadds, MKmisc, stats, tableone, and VIM, and IBM SPSS version 22 (IBM Corp, Armonk, NY).

Results

Recruitment

During the 40-month recruitment period, 3928 patients were screened for eligibility, with a total of 239 (6.08% (239/3928)

of all screened) randomized. Of these, 10.9% (26/239) were included based on HADS-D only, 38.1% (91/239) based on HADS-A only, and 51.0% (122/239) based on both subscales. The main reasons for exclusion were the following: being unable/unwilling to use the internet or mobile phone, followed by scoring <8 on both HADS subscales and language difficulties. In total, 34.6% (1359/3928) declined participation or did not return the informed consent form. Follow-up assessment was completed by 88.3% (211/239) of all patients, with a significantly higher percentage of completers in the control group (94.3%, 115/122) compared with the treatment group (82.1%, 96/117; Pearson χ^2_1=8.6, *P*=.003). See Figure 1 for a study flowchart.

Patient Characteristics

Baseline patient characteristics were similar in both groups (Table 2). On average, patients were 59.6 years of age (SD 8.49), 33.5% (80/239) were women, 41.8% (100/239) had university level of education, 60.3% (144/239) were employed, and 18.0% (43/239) were taking antidepressant and/or anxiolytic medication in both groups. The corresponding percentages after treatment was 19.6% (18/92; 25 missing values) in the treatment group and 15.6% (18/115; 7 missing values) in the control group (*P*=.37).

Figure 1. Flowchart of patients through the U-CARE Heart trial. iCBT: Internet-based cognitive behavioral therapy; HADS-D: Hospital Anxiety and Depression Scale-Depression subscale; HADS-A: Hospital Anxiety and Depression Scale-Anxiety subscale.

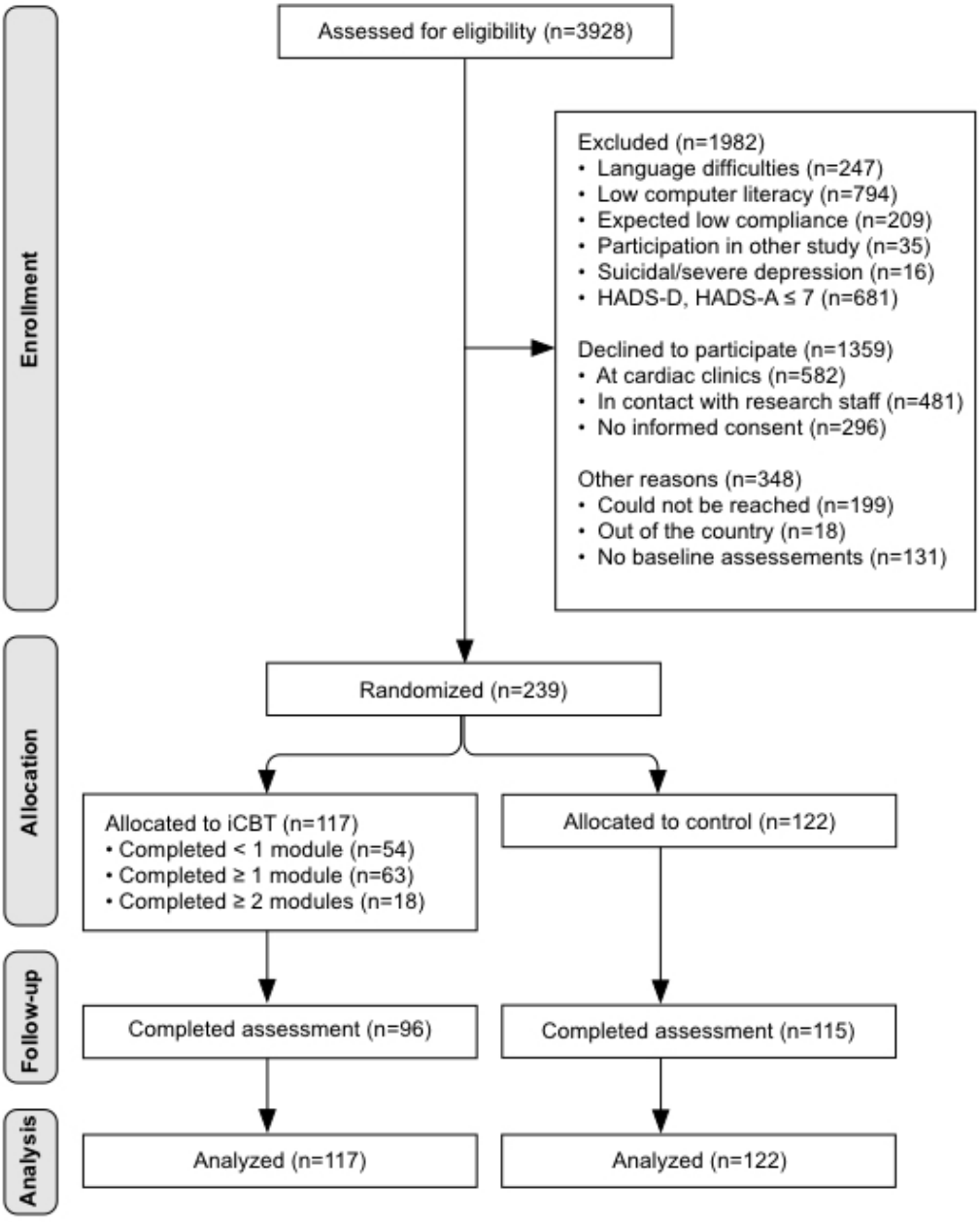


Table 2. Patient characteristics. Observed data (no imputations).

Characteristics	iCBT[a] (n=117)	TAU[b] (n=122)	Missing n (%)
Sociodemographic			
Age in years, mean (SD)	58.4 (9.0)	60.8 (7.8)	
Women, n (%)	44 (37.6)	36 (29.5)	
Occupation, n (%)			14 (5.8)
Employed	78 (66.7)	66 (54.1)	
Unemployed	4 (3.4)	2 (1.6)	
Retired	33 (28.2)	37 (30.3)	
Sick leave	2 (1.7)	1 (0.8)	
Other	0 (0.0)	2 (1.6)	
Highest level of education, n (%)			
Elementary	22 (18.8)	26 (21.3)	
High school	45 (38.5)	46 (37.7)	
University	50 (42.7)	50 (41.0)	
In a relationship, n (%)	99 (84.6)	101 (82.8)	
Children in the household, n (%)	43 (36.8)	34 (27.9)	
Country of birth other than Sweden, n (%)	21 (17.9)	15 (12.3)	
Smoking, n (%)	6 (5.1)	8 (6.6)	
Alcohol, standard drinks/week (SD)	5.7 (13.7)	5.5 (6.1)	
Leisure time physical activity, n (%)			
High activity	19 (16.2)	24 (19.7)	
Moderate activity	52 (44.4)	65 (53.3)	
Low activity	37 (31.6)	26 (21.3)	
Sedentary lifestyle	9 (7.7)	7 (5.7)	
Psychotropic medicine, n (%)			
Anxiolytics	10 (8.5)	7 (5.7)	
Antidepressants	11 (9.4)	15 (12.3)	
No	98 (83.8)	102 (83.6)	
Other current counseling, n (%)	30 (25.7)	28 (22.9)	
Medical history			
Myocardial infarction, n (%)	19 (16.2)	13 (10.7)	10 (4.2)
Diabetes, n (%)	21 (17.9)	19 (15.6)	9 (3.7)
Hypertension, n (%)	42 (35.9)	51 (41.8)	9 (3.7)
Hyperlipidemia, n (%)	26 (22.2)	27 (22.1)	9 (3.7)
Stroke, n (%)	0 (0.0)	4 (3.3)	4 (1.7)
Heart failure, n (%)	4 (3.4)	2 (1.6)	16 (6.7)
Cardiac status and medication			
Any angina/chest pain, n (%)	34 (29.0)	32 (26.2)	30 (12.6)
Blood pressure <140/90, n (%)	66 (56.4)	78 (63.9)	31 (13.0)
Body mass index, mean (SD), kg/m^2	27.8 (5.0)	27.4 (4.0)	18 (7.5)
Beta-blockers at discharge, n (%)	104 (88.9)	106 (86.9)	9 (3.8)
Statins at discharge, n (%)	110 (94.0)	115 (94.3)	9 (3.8)

Characteristics	iCBT[a] (n=117)	TAU[b] (n=122)	Missing n (%)
ACE[c] inhibitor/ARB[d] at discharge, n (%)	89 (76.1)	96 (78.7)	9 (3.8)
DAPT[e] at discharge, n (%)	107 (91.4)	107 (87.7)	10 (4.2)

[a]iCBT: internet-based cognitive behavioral therapy.

[b]TAU: treatment as usual.

[c]ACE:angiotensin-converting enzyme.

[d]ARB: angiotensin receptor blocker.

[e]DAPT: dual antiplatelet therapy.

In total, 33.1% (79/239) were sedentary or reported low levels of exercise, 16.7% (40/239) had previous diabetes mellitus, and 13.4% (32/239) had a previous MI. At baseline, 25.7% (30/117) in the iCBT group and 22.9% (28/122) in the control group had regular contact with a counselor within TAU. The corresponding percentage at follow-up was 21.1% (19/90; 27 missing values) in the iCBT group and 27.2% (31/114; 8 missing values) in the control group (P=.33).

Primary Outcomes

There was no difference in HADS-T scores at baseline between the iCBT and the control group (t_{237}=0.56, P=.85). There was a general reduction in HADS-T over time in the total study sample (mean delta=−5.1; t_{237}=12.92, P<.001).

The main analysis showed no effect of treatment on HADS-T at follow-up (beta=−0.47, 95% CI −1.95 to 1.00, P=.53). Furthermore, the main analysis showed that men scored lower on HADS-T compared with women at follow-up (beta=−2.04, 95% CI −3.60 to −0.47], P=.01), and there was a borderline significant reduction in HADS-T per unit increase in age (beta=−0.08, 95% CI −0.16 to 0.01, P=.09) at follow-up. There was no interaction between treatment and sex, or treatment and age, on HADS-T (P for both >0.19). Congruent with the main analysis, separate exploratory analyses showed no effect of treatment on either HADS-A or HADS-D subscales (Table 3).

Results of the sensitivity analyses were consistent with the ITT analysis. Both the PP analysis (beta=−0.87, 95% CI −2.47 to 0.72, P=.28) and the analysis with observed data (beta=−0.55, 95% CI −2.04 to 0.93], P=.46) with HADS-T as the outcome yielded no effect of treatment (Table 3).

Secondary Outcomes

Additional multiple linear models showed no effect of treatment on the secondary outcomes MADRS-S, CAQ, or BADS-SF at follow-up (Table 3).

Adverse Events

Two patients in the iCBT group and 3 patients in the control group reported severe depression (MADRS-S>34) or suicidal ideation (MADRS-S item 9 >3) at follow-up.

Adherence

Treatment adherence was low, with 46.2% (54/117) of the iCBT group not completing the introductory module, 38.4% (45/117) completing the introductory module only, and 15.4% (18/117) completing additional modules (Figure 2). Furthermore, only 0.9% (1/117) adhered to the treatment [26] by completing the recommended number of 14 steps within the 14-week treatment period. The number of completed homework assignments was not associated with change in HADS-T at follow-up, r_s=.07, P=.53.

Table 3. Outcomes at baseline and follow-up, change scores, and treatment effects. Mean (SD) and change are calculated from observed data. Effect estimates (beta) are pooled adjusted coefficients for treatment (internet-based cognitive behavioral therapy, iCBT) versus control (treatment as usual) on follow-up outcomes adjusted for sex, age, and baseline levels of the respective outcomes after multiple imputation.

Outcome	Baseline, mean (SD)	Follow-up, mean (SD)	Change	Effect, Beta (95% CI)	*P* value
HADS-T[a]				−.47 (−1.95 to 1.00)	.53
iCBT	18.3 (4.9)	12.8 (5.9)	−5.5		
Control	18.6 (5.0)	13.6 (6.8)	−5.0		
HADS-A[b]				−.09 (−0.91 to 0.72)	.82
iCBT	10.9 (2.4)	7.4 (3.2)	−3.5		
Control	10.8 (2.5)	7.3 (3.7)	−3.5		
HADS-D[c]				−.45 (−1.34 to 0.44)	.32
iCBT	9.9 (2.2)	6.6 (3.3)	−3.3		
Control	10.3 (2.5)	8.0 (3.8)	−2.3		
MADRS-S[d]				−.58 (−2.20 to 1.04)	.48
iCBT	14.8 (6.4)	12.0 (7.2)	−2.8		
Control	15.9 (7.2)	13.3 (7.6)	−2.6		
CAQ[e]				−.73 (−2.83 to 1.38)	.50
iCBT	26.1 (10.3)	21.5 (10.2)	−5.4		
Control	25.3 (10.8)	22.0 (11.4)	−3.3		
BADS-SF[f]				−.50 (−2.31 to 1.30)	.58
iCBT	21.2 (6.1)	21.4 (6.9)	0.2		
Control	21.4 (7.7)	21.6 (7.2)	0.2		

[a]HADS-T: Hospital Anxiety and Depression Scale total score.

[b]HADS-A: Hospital Anxiety and Depression Scale anxiety subscale.

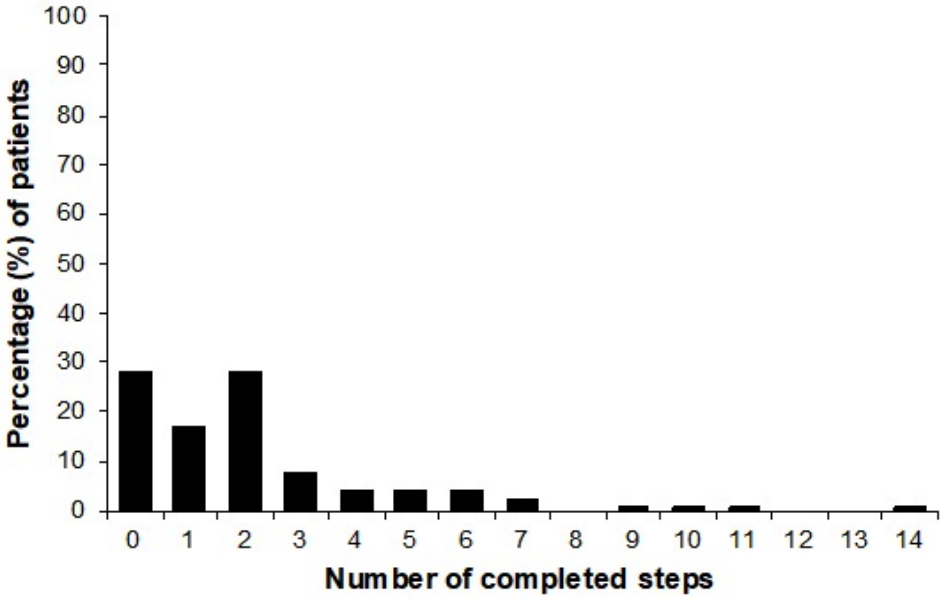
[c]HADS-D: Hospital Anxiety and Depression Scale depression subscale.

[d]MADRS-S: The Montgomery-Asberg Depression Rating Scale-Self Rated.

[e]CAQ: Cardiac Anxiety Questionnaire.

[f]BADS-SF: Behavioral Activation for Depression Scale-Short Form.

Figure 2. Proportion of patients completing different number of steps in the internet-based cognitive behavioral therapy.



Discussion

Principal Findings

In this RCT, we evaluated the effectiveness of a therapist-guided, tailored iCBT treatment compared with TAU to reduce symptoms of depression and anxiety among recent MI patients. Both groups reported a decreased level of symptoms of depression and anxiety over time to a similar extent, with no difference between groups at follow-up. Adherence was low compared with other tailored iCBT interventions for depression and anxiety [26], indicating most patients allocated to iCBT received only a small treatment dose.

Overall, 6.08% (239/3928) of the screened patients were randomized. The main reasons for exclusion were reported as being unable or unwilling to use internet or mobile phone, HADS score below the inclusion threshold (<8), and language difficulties. Furthermore, a substantial number of patients screened for eligibility declined to participate. Reasons for declining are not fully known, but might include low perceived need for help or a preference for other treatment alternatives. Low interest in iCBT treatment among cardiac patients has also been reported previously. The InterHerz study [31], which resembles the U-CARE Heart RCT, ended prematurely because of low recruitment rates (12 patients in 6 months; personal communication October 1, 2017 with Professor Nadine Messerli-Bürgy). Negative attitudes toward, and low intentions to use, internet-based psychological interventions have been reported previously in other populations [32,33]. Access to face-to-face counseling and psychotropic medicine is readily available and is of good quality in standard MI care in Sweden, with an estimated 95% of cardiac clinics in Sweden assessing and referring patients with mental health difficulties to appropriate care [34]. As such, a low interest in iCBT interventions among cardiac patients may be expected, which in turn may be a barrier, or at least a challenge, for implementation in routine care.

Previous findings suggest iCBT as an effective treatment for comorbid symptoms of depression and anxiety in patients with somatic conditions [13], and effective psychological interventions for emotional distress related to coronary heart disease are characterized by CBT techniques [9]. Given this, the lack of effect of the iCBT intervention found in this study may have several explanations. Two factors might be low treatment adherence in the intervention group and a significant spontaneous improvement in the control group. Most iCBT studies with positive results are efficacy studies based on self-referral by people seeking help on the internet. In this study, we recruited patients within a routine care setting using screening methods. However, treatment adherence to iCBT interventions has been found to be lower in effectiveness studies in primary care samples compared with samples recruited from Web-based self-referral [35,36]. It is likely that patients actively seeking out and self-referring to iCBT are more prone to stay active in treatment compared with those who are screened and offered participation. Indeed, some of the patients in this study reported that their strongest reason for joining the study was to assist in research rather than seeking help for their depression or anxiety. In addition, reporting severe depressive symptoms was an exclusion criterion.

Another important factor that may have influenced treatment adherence was related to iCBT characteristics. The treatment and U-CARE portal used to deliver it were developed in consultation with patients with personal experience of depression and anxiety post-MI. In spite of this effort, the content and design of the intervention might not have been adjusted enough to end users' needs, for example, in terms of relevance and workload. Indeed, treatment burden and failure to tailor content adequately are associated with negative iCBT user experience [37]. Moreover, therapist support has been shown to significantly improve iCBT treatment adherence and effect [38]. However, the amount of support needed in different populations may vary, with some patient populations potentially benefitting from more extensive support. Indeed, extending iCBT support through additional weekly telephone calls has been found to improve treatment adherence [39]. As such, real-time therapist support via telephone might have helped patients engage with and adhere to treatment over time. Furthermore, some previous successful psychological interventions (with high adherence) for cardiac patients [40-42] have been group-based CBT with a process-oriented focus. It remains to be investigated in a randomized trial if process-oriented, group-based formats are necessary in psychological interventions for cardiac patients.

Patient characteristics may affect treatment adherence. The mean age of patients in this study was >10 years lower than the average MI population, but higher compared with other iCBT studies of patients with depression and anxiety [35,43,44]. Older age is correlated with lower computer literacy [45]. It is possible that patients experiencing technological difficulties were less active in treatment. Furthermore, the level of education was somewhat lower compared with other iCBT studies (40% university level vs 50-60%) [35,43,44], a factor further associated with low adherence to psychological treatment [46].

Both groups reported improved psychological symptoms over time, with regression to the mean potentially explaining this pattern. In addition, a substantial spontaneous improvement has been reported for MI patients in symptoms of both anxiety and depression over time [47]. Our patients were recruited about 10 weeks post-MI to avoid spontaneous recovery diluting any treatment effects. However, this recruitment strategy may have resulted in patients finding other ways to improve their psychological well-being. Moreover, more patients in the control group than in the iCBT group reported initiating a contact with a local counselor during the study period, but the difference was not significant.

Strengths and Limitations

This trial recruited patients from 25 hospitals in both rural and urban areas in Sweden. The content and design of the portal and the treatment were developed in consultation with patients with personal experience of emotional distress post-MI to increase acceptability, relevance, and usability. We prepared a detailed statistical analysis plan and prespecified adjustment by covariates to ensure a transparent analysis procedure [48]. We have provided detailed descriptions of the intervention and its delivery, in line with recent reporting guidelines [49], enabling

comparison with other iCBT treatments targeting cardiac patients. Therapist support was provided by licensed psychologists, specialized in the CBT methodology. Despite all efforts to develop a user-friendly and relevant iCBT treatment, adherence to treatment was low. Given the obtained dose of treatment was low, the effect of the treatment might be difficult to evaluate. The PP analysis did not differ from the ITT analysis. However, our definition of PP may be criticized of being too liberal (completion of only 1 homework assignment). Moreover, HADS was developed as a screening measure and might not be sensitive enough to detect minor changes over time. However, none of the more sensitive secondary outcomes assessments indicated an effect. Diagnostic interviews might have been a more valid assessment of symptoms of depression and anxiety. Furthermore, the initial cut-off of >10 in any of the HADS subscales was lowered early in the study to >7 to increase recruitment rate. Patients reporting a low level of depression and anxiety have less room for improvement, resulting in a reduced likelihood of detecting a treatment effect [50]. This might also have resulted in inclusion of patients experiencing a low level of emotional distress, and consequently low perceived need for psychological help.

Future Directions

Effective and accessible psychological treatments are important, given symptoms of depression and anxiety are common post-MI. Despite the success of iCBT trials [13], using self-referral recruitment methods for patients with a range of comorbid physical and mental conditions, alongside calls for the widespread implementation of eHealth interventions for cardiac populations [10], this study questions the promise of iCBT for MI patients recruited at cardiac clinics. As such, future research should examine the potential differences in terms of iCBT acceptability between populations recruited via self-referral versus clinical settings. Furthermore, increased efforts are needed to better understand how to improve treatment adherence. Such efforts may include exploratory studies investigating factors related to treatment acceptability. In interviews with participants in this study, some challenges have been identified. These are described elsewhere [27]. Finally, our results support the notion that systematic development and feasibility testing, in close collaboration with potential end users, should be undertaken to improve treatment relevance and acceptability [51]. Although this study was preceded by both semistructured feasibility testing and an internal pilot study, this was apparently not sufficient.

Conclusions

In a randomized trial, we evaluated the effects of a therapist-guided, tailored iCBT intervention for depression and anxiety versus TAU among recent MI. Both groups reported less emotional distress after treatment, but iCBT did not significantly reduce symptoms of depression or anxiety in comparison with TAU. This lack of difference in treatment outcome may be explained by low treatment adherence, with further investigation into reasons for poor treatment adherence warranted. This study suggests that further research is required into the acceptability and feasibility of iCBT for an MI population before wide-scale implementation of similar eHealth solutions for this patient group.

Acknowledgments

This study is part of U-CARE, which is a Strategic Research environment funded by the Swedish Research Council (dnr 2009–1093). U-CARE Heart is also funded by the Swedish Heart and Lung Association (dnr E 148/11); the Uppsala-Örebro Regional Research Council (dnr 22 86 61); Swedish Research Council for Health, Working Life, and Welfare (dnr 2014–4947); and the Vårdal foundation (dnr 2014–0114). The authors would like to thank all patients who participated in this study. The authors would also like to thank all participating hospitals, collaborating nurses, and doctors who helped them in the recruitment process. The authors are also grateful to research assistants Malin Fridlund and Kata Bordas who helped out in recruiting and informing patients in the study. The authors thank Amanda Romppala for helping out administratively. The authors would also like to acknowledge Maria Bertilsson, Lars Lindhagen, and Ronnie Pingel at Uppsala University for statistical support.

Authors' Contributions

CH, EO, GB, LvE, and FN designed the study. JW, FN, EW, and EO analyzed the data. FN and EW drafted the manuscript. All authors critically revised and edited the draft and approved of the final version. JW, FN, EW, EO, and CH had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Sitemap of the Web-based portal.

[JPG File, 8MB - jmir_v20i3e88_app1.jpg]

Multimedia Appendix 2

Screenshot of the U-CARE Heart portal.

[JPG File, 2MB - jmir_v20i3e88_app2.jpg]

Multimedia Appendix 3

CONSORT E-HEALTH checklist (v 1.6.1).

[PDF File (Adobe PDF File), 852KB - jmir_v20i3e88_app3.pdf]

References

1. Thombs BD, Bass EB, Ford DE, Stewart KJ, Tsilidis KK, Patel U, et al. Prevalence of depression in survivors of acute myocardial infarction. J Gen Intern Med 2006 Jan;21(1):30-38 [FREE Full text] [doi: 10.1111/j.1525-1497.2005.00269.x] [Medline: 16423120]
2. Roest AM, Martens EJ, Denollet J, de Jonge P. Prognostic association of anxiety post myocardial infarction with mortality and new cardiac events: a meta-analysis. Psychosom Med 2010 Jul;72(6):563-569. [doi: 10.1097/PSY.0b013e3181dbff97] [Medline: 20410247]
3. Lane D, Carroll D, Ring C, Beevers DG, Lip GY. The prevalence and persistence of depression and anxiety following myocardial infarction. Br J Health Psychol 2002 Feb;7(Pt 1):11-21. [doi: 10.1348/135910702169321] [Medline: 14596714]
4. Meijer A, Conradi HJ, Bos EH, Thombs BD, van Melle JP, de Jonge P. Prognostic association of depression following myocardial infarction with mortality and cardiovascular events: a meta-analysis of 25 years of research. Gen Hosp Psychiatry 2011;33(3):203-216 [FREE Full text] [doi: 10.1016/j.genhosppsych.2011.02.007] [Medline: 21601716]
5. Benyamini Y, Roziner I, Goldbourt U, Drory Y, Gerber Y, Israel Study Group on First Acute Myocardial Infarction. Depression and anxiety following myocardial infarction and their inverse associations with future health behaviors and quality of life. Ann Behav Med 2013 Dec;46(3):310-321. [doi: 10.1007/s12160-013-9509-3] [Medline: 23645421]
6. Davidson KW, Rieckmann N, Clemow L, Schwartz JE, Shimbo D, Medina V, et al. Enhanced depression care for patients with acute coronary syndrome and persistent depressive symptoms: coronary psychosocial evaluation studies randomized controlled trial. Arch Intern Med 2010 Apr 12;170(7):600-608 [FREE Full text] [doi: 10.1001/archinternmed.2010.29] [Medline: 20386003]
7. Huffman JC, Mastromauro CA, Beach SR, Celano CM, DuBois CM, Healy BC, et al. Collaborative care for depression and anxiety disorders in patients with recent cardiac events: the Management of Sadness and Anxiety in Cardiology (MOSAIC) randomized clinical trial. JAMA Intern Med 2014 Jun;174(6):927-935. [doi: 10.1001/jamainternmed.2014.739] [Medline: 24733277]
8. Richards SH, Anderson L, Jenkinson CE, Whalley B, Rees K, Davies P, et al. Psychological interventions for coronary heart disease. Cochrane Database Syst Rev 2017 Dec 28;4:CD002902. [doi: 10.1002/14651858.CD002902.pub4] [Medline: 28452408]
9. Dickens C, Cherrington A, Adeyemi I, Roughley K, Bower P, Garrett C, et al. Characteristics of psychological interventions that improve depression in people with coronary heart disease: a systematic review and meta-regression. Psychosom Med 2013 Feb;75(2):211-221. [doi: 10.1097/PSY.0b013e31827ac009] [Medline: 23324874]
10. Cowie MR, Chronaki CE, Vardas P. e-Health innovation: time for engagement with the cardiology community. Eur Heart J 2013 Jul;34(25):1864-1868. [doi: 10.1093/eurheartj/ehs153] [Medline: 22733834]
11. Rebello TJ, Marques A, Gureje O, Pike KM. Innovative strategies for closing the mental health treatment gap globally. Curr Opin Psychiatry 2014 Jul;27(4):308-314. [doi: 10.1097/YCO.0000000000000068] [Medline: 24840160]
12. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
13. van Beugen S, Ferwerda M, Hoeve D, Rovers MM, Spillekom-van Koulil S, van Middendorp H, et al. Internet-based cognitive behavioral therapy for patients with chronic somatic conditions: a meta-analytic review. J Med Internet Res 2014;16(3):e88 [FREE Full text] [doi: 10.2196/jmir.2777] [Medline: 24675372]
14. Glozier N, Christensen H, Naismith S, Cockayne N, Donkin L, Neal B, et al. Internet-delivered cognitive behavioural therapy for adults with mild to moderate depression and high cardiovascular disease risks: a randomised attention-controlled trial. PLoS One 2013;8(3):e59139 [FREE Full text] [doi: 10.1371/journal.pone.0059139] [Medline: 23555624]
15. Norlund F, Olsson EM, Burell G, Wallin E, Held C. Treatment of depression and anxiety with internet-based cognitive behavior therapy in patients with a recent myocardial infarction (U-CARE Heart): study protocol for a randomized controlled trial. Trials 2015 Apr 11;16:154 [FREE Full text] [doi: 10.1186/s13063-015-0689-y] [Medline: 25873137]
16. Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x]
17. Svanborg P, Asberg M. A comparison between the Beck Depression Inventory (BDI) and the self-rating version of the Montgomery Asberg Depression Rating Scale (MADRS). J Affect Disord 2001 May;64(2-3):203-216. [Medline: 11313087]
18. Andersson G, Estling F, Jakobsson E, Cuijpers P, Carlbring P. Can the patient decide which modules to endorse? An open trial of tailored internet treatment of anxiety disorders. Cogn Behav Ther 2011;40(1):57-64. [doi: 10.1080/16506073.2010.529457] [Medline: 21337215]

19. Paxling B, Lundgren S, Norman A, Almlöv J, Carlbring P, Cuijpers P, et al. Therapist behaviours in internet-delivered cognitive behaviour therapy: analyses of e-mail correspondence in the treatment of generalized anxiety disorder. Behav Cogn Psychother 2013 May;41(3):280-289. [doi: 10.1017/S1352465812000240] [Medline: 22717145]
20. Jernberg T, Attebring MF, Hambraeus K, Ivert T, James S, Jeppsson A, et al. The Swedish Web-system for enhancement and development of evidence-based care in heart disease evaluated according to recommended therapies (SWEDEHEART). Heart 2010 Oct;96(20):1617-1621. [doi: 10.1136/hrt.2010.198804] [Medline: 20801780]
21. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. J Psychosom Res 2002 Feb;52(2):69-77. [Medline: 11832252]
22. van Ballegooijen W, Riper H, Cuijpers P, van Oppen P, Smit JH. Validation of online psychometric instruments for common mental health disorders: a systematic review. BMC Psychiatry 2016;16:45 [FREE Full text] [doi: 10.1186/s12888-016-0735-7] [Medline: 26915661]
23. Holländare F, Andersson G, Engström I. A comparison of psychometric properties between internet and paper versions of two depression instruments (BDI-II and MADRS-S) administered to clinic patients. J Med Internet Res 2010;12(5):e49 [FREE Full text] [doi: 10.2196/jmir.1392] [Medline: 21169165]
24. Manos RC, Kanter JW, Luo W. The behavioral activation for depression scale-short form: development and validation. Behav Ther 2011 Dec;42(4):726-739. [doi: 10.1016/j.beth.2011.04.004] [Medline: 22036000]
25. Eifert GH, Thompson RN, Zvolensky MJ, Edwards K, Frazer NL, Haddad JW, et al. The Cardiac Anxiety Questionnaire: development and preliminary validity. Behav Res Ther 2000 Oct;38(10):1039-1053. [Medline: 11004742]
26. Păsărelu CR, Andersson G, Bergman Nordgren L, Dobrean A. Internet-delivered transdiagnostic and tailored cognitive behavioral therapy for anxiety and depression: a systematic review and meta-analysis of randomized controlled trials. Cogn Behav Ther 2017 Jan;46(1):1-28. [doi: 10.1080/16506073.2016.1231219] [Medline: 27712544]
27. Wallin E, Norlund F, Olsson EM, Burell G, Held C, Carlsson T. Acceptability of internet-based cognitive behavioral therapy for adults with depression and anxiety after a myocardial infarction: a mixed methods study. J Med Internet Res 2018;In Press. [doi: 10.2196/jmir.9690]
28. van Buuren S, Groothuis-Oudshoorn K. mice: multivariate imputation by chained equations in R. J Stat Softw 2011;45(3). [doi: 10.18637/jss.v045.i03]
29. Rubin D. Multiple Imputation for Nonresponse in Surveys. New York: Wiley & Sons; 1987.
30. R Core Team, R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria, Austria: R Foundation for Statistical Computing; 2015.
31. Messerli-Bürgy N, Barth J, Berger T. The InterHerz project--a web-based psychological treatment for cardiac patients with depression: study protocol of a randomized controlled trial. Trials 2012;13:245 [FREE Full text] [doi: 10.1186/1745-6215-13-245] [Medline: 23273042]
32. Musiat P, Goldstone P, Tarrier N. Understanding the acceptability of e-mental health--attitudes and expectations towards computerised self-help treatments for mental health problems. BMC Psychiatry 2014;14:109 [FREE Full text] [doi: 10.1186/1471-244X-14-109] [Medline: 24725765]
33. Wallin EE, Mattsson S, Olsson EM. The preference for internet-based psychological interventions by individuals without past or current use of mental health treatment delivered online: a survey study with mixed-methods analysis. JMIR Ment Health 2016;3(2):e25 [FREE Full text] [doi: 10.2196/mental.5324] [Medline: 27302200]
34. Ogmundsdottir Michelsen H, Hagstrom E, Sjolin I, Schlyter M, Kiessling MA, Held C, et al. Swedish cardiac rehabilitation programmes; a descriptive nationwide analysis - the perfect CR study. Eur Heart J (Supplement 1) 2017;38:1032-1032. [doi: 10.1093/eurheartj/ehx493.P4912]
35. Newby JM, Mackenzie A, Williams AD, McIntyre K, Watts S, Wong N, et al. Internet cognitive behavioural therapy for mixed anxiety and depression: a randomized controlled trial and evidence of effectiveness in primary care. Psychol Med 2013 Dec;43(12):2635-2648. [doi: 10.1017/S0033291713000111] [Medline: 23419552]
36. Allen AR, Newby JM, Mackenzie A, Smith J, Boulton M, Loughnan SA, et al. Internet cognitive-behavioural treatment for panic disorder: randomised controlled trial and evidence of effectiveness in primary care. BJPsych Open 2016 Mar;2(2):154-162 [FREE Full text] [doi: 10.1192/bjpo.bp.115.001826] [Medline: 27703768]
37. Knowles SE, Lovell K, Bower P, Gilbody S, Littlewood E, Lester H. Patient experience of computerised therapy for depression in primary care. Br Med J Open 2015 Nov 30;5(11):e008581. [doi: 10.1136/bmjopen-2015-008581] [Medline: 26621513]
38. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
39. O'Mahen HA, Richards DA, Woodford J, Wilkinson E, McGinley J, Taylor RS, et al. Netmums: a phase II randomized controlled trial of a guided Internet behavioural activation treatment for postpartum depression. Psychol Med 2014 Jun;44(8):1675-1689 [FREE Full text] [doi: 10.1017/S0033291713002092] [Medline: 24148703]
40. Orth-Gomér K, Schneiderman N, Wang H, Walldin C, Blom M, Jernberg T. Stress reduction prolongs life in women with coronary disease: the Stockholm Women's Intervention Trial for Coronary Heart Disease (SWITCHD). Circ Cardiovasc Qual Outcomes 2009 Jan;2(1):25-32 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.108.812859] [Medline: 20031809]

41. Gulliksson M, Burell G, Vessby B, Lundin L, Toss H, Svärdsudd K. Randomized controlled trial of cognitive behavioral therapy vs standard treatment to prevent recurrent cardiovascular events in patients with coronary heart disease: Secondary Prevention in Uppsala Primary Health Care project (SUPRIM). Arch Intern Med 2011 Jan 24;171(2):134-140. [doi: 10.1001/archinternmed.2010.510] [Medline: 21263103]
42. Blumenthal JA, Sherwood A, Smith PJ, Watkins L, Mabe S, Kraus WE, et al. Enhancing cardiac rehabilitation with stress management training: a randomized, clinical efficacy trial. Circulation 2016 Apr 05;133(14):1341-1350 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.018926] [Medline: 27045127]
43. Hedman E, Ljótsson B, Kaldo V, Hesser H, El AS, Kraepelien M, et al. Effectiveness of Internet-based cognitive behaviour therapy for depression in routine psychiatric care. J Affect Disord 2014 Feb;155:49-58. [doi: 10.1016/j.jad.2013.10.023] [Medline: 24238951]
44. Lindner P, Nyström MB, Hassmén P, Andersson G, Carlbring P. Who seeks ICBT for depression and how do they get there? Effects of recruitment source on patient demographics and clinical characteristics. Internet Interv 2015 May;2(2):221-225. [doi: 10.1016/j.invent.2015.04.002]
45. Crabb RM, Cavanagh K, Proudfoot J, Learmonth D, Rafie S, Weingardt KR. Is computerized cognitive-behavioural therapy a treatment option for depression in late-life? A systematic review. Br J Clin Psychol 2012 Nov;51(4):459-464. [doi: 10.1111/j.2044-8260.2012.02038.x] [Medline: 23078214]
46. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
47. Hanssen TA, Nordrehaug JE, Eide GE, Bjelland I, Rokne B. Anxiety and depression after acute myocardial infarction: an 18-month follow-up study with repeated measures and comparison with a reference population. Eur J Cardiovasc Prev Rehabil 2009 Dec;16(6):651-659. [doi: 10.1097/HJR.0b013e32832e4206] [Medline: 19707149]
48. [No authors listed]. ICH harmonised tripartite guideline. Statistical principles for clinical trials. International Conference on Harmonisation E9 Expert Working Group. Stat Med 1999 Aug 15;18(15):1905-1942. [Medline: 10532877]
49. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. Br Med J 2014;348:g1687 [FREE Full text] [Medline: 24609605]
50. Bower P, Kontopantelis E, Sutton A, Kendrick T, Richards DA, Gilbody S, et al. Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. Br Med J 2013;346:f540 [FREE Full text] [Medline: 23444423]
51. Ennis L, Wykes T. Impact of patient involvement in mental health research: longitudinal study. Br J Psychiatry 2013 Nov;203(5):381-386 [FREE Full text] [doi: 10.1192/bjp.bp.112.119818] [Medline: 24029538]

Abbreviations

BADS-SF: Behavioral Activation for Depression Scale-Short Form
CAQ: Cardiac Anxiety Questionnaire
CBT: cognitive behavioral therapy
HADS: Hospital Anxiety and Depression Scale
HADS-A: HADS-anxiety score
HADS-D: HADS-depression score
HADS-T: HADS-total score
iCBT: internet-based cognitive behavioral therapy
ITT: intention-to-treat
MADRS-S: Montgomery-Asberg Depression Rating Scale-Self Rated
MI: myocardial infarction
PP: per protocol
RCT: randomized controlled trial
SMS: short message service
TAU: treatment as usual

Edited by G Eysenbach; submitted 22.12.17; peer-reviewed by A Conway, A Parks; comments to author 18.01.18; revised version received 02.02.18; accepted 08.02.18; published 08.03.18.

Please cite as:
Norlund F, Wallin E, Olsson EMG, Wallert J, Burell G, von Essen L, Held C
Internet-Based Cognitive Behavioral Therapy for Symptoms of Depression and Anxiety Among Patients With a Recent Myocardial Infarction: The U-CARE Heart Randomized Controlled Trial
J Med Internet Res 2018;20(3):e88
URL: http://www.jmir.org/2018/3/e88/
doi:10.2196/jmir.9710
PMID:29519777

Original Paper

Detection of Cases of Noncompliance to Drug Treatment in Patient Forum Posts: Topic Model Approach

Redhouane Abdellaoui[1], MSc; Pierre Foulquié[2], MSc; Nathalie Texier[2], PharmD; Carole Faviez[2], MSc; Anita Burgun[1,3], MD, PhD; Stéphane Schück[2], MSc, MD

[1]Unité de Mixte de Recherche 1138 Team 22, Institut National de la Santé et de la Recherche Médicale / Université Pierre et Marie Curie, Paris, France
[2]Kappa Santé, Innovation (Kap Code), Paris, France
[3]Medical Informatics, Hôpital Européen Georges-Pompidou, Assistance Publique-Hôpitaux de Paris, Paris, France

Corresponding Author:
Redhouane Abdellaoui, MSc
Unité de Mixte de Recherche 1138 Team 22
Institut National de la Santé et de la Recherche Médicale / Université Pierre et Marie Curie
15 Rue de l'École de Médecine
Paris, 75006
France
Phone: 33 648094269
Email: redhouane.a@gmail.com

Abstract

Background: Medication nonadherence is a major impediment to the management of many health conditions. A better understanding of the factors underlying noncompliance to treatment may help health professionals to address it. Patients use peer-to-peer virtual communities and social media to share their experiences regarding their treatments and diseases. Using topic models makes it possible to model themes present in a collection of posts, thus to identify cases of noncompliance.

Objective: The aim of this study was to detect messages describing patients' noncompliant behaviors associated with a drug of interest. Thus, the objective was the clustering of posts featuring a homogeneous vocabulary related to nonadherent attitudes.

Methods: We focused on escitalopram and aripiprazole used to treat depression and psychotic conditions, respectively. We implemented a probabilistic topic model to identify the topics that occurred in a corpus of messages mentioning these drugs, posted from 2004 to 2013 on three of the most popular French forums. Data were collected using a Web crawler designed by Kappa Santé as part of the Detec't project to analyze social media for drug safety. Several topics were related to noncompliance to treatment.

Results: Starting from a corpus of 3650 posts related to an antidepressant drug (escitalopram) and 2164 posts related to an antipsychotic drug (aripiprazole), the use of latent Dirichlet allocation allowed us to model several themes, including interruptions of treatment and changes in dosage. The topic model approach detected cases of noncompliance behaviors with a recall of 98.5% (272/276) and a precision of 32.6% (272/844).

Conclusions: Topic models enabled us to explore patients' discussions on community websites and to identify posts related with noncompliant behaviors. After a manual review of the messages in the noncompliance topics, we found that noncompliance to treatment was present in 6.17% (276/4469) of the posts.

(J Med Internet Res 2018;20(3):e85) doi:10.2196/jmir.9222

KEYWORDS

medication adherence; compliance; infodemiology; social media; text mining; depression; psychosis; peer-to-peer support; virtual community

Introduction

Background

A report published by the World Health Organization (WHO) in 2003 highlighted that noncompliance (or nonadherence) to long-term treatment was a worldwide problem detrimental to the overall effectiveness of the health system [1]. Compliance is defined in this report as the degree of correspondence between a patient's behavior (taking medications, following hygiene rules, and diet) and the recommendations made by a health care professional (HCP). Noncompliance with these recommendations has an impact on patients' quality of life (QoL), outcomes, and health costs.

The WHO identified several causes of nonadherence to therapies, including the characteristics of the health system, the patient's disease, and the course of treatment. For patients with depression, observance is linked to the frequency of administration of a drug and to concomitant therapy. For patients suffering from cancer, the fear of adverse effects (AEs) related to the treatment has negative impact on adherence. For diabetic patients, adherence may vary with age, sex, and the relationship with the physician. Several meta-analyses showed that current methods of improving medication adherence for chronic diseases were mostly complex and not very effective [2,3]. The Cochrane group concluded that (1) means to measure adherence more systematically and objectively and (2) innovations to assist patients to follow medication prescriptions for long-term medical disorders were major points to be considered in that field. Considering social media as platforms where patients can discuss about their treatments and share testimonies, they could be a new data source to measure adherence to treatment.

The use of social media allows large groups of people to create and share information, opinions, and experiences about health conditions and medications through discussions [4]. Social media provide pharmacovigilance experts with a relevant source of information [5]. The example of benfluorex [6] illustrated how social media could be valuable sources for experts. Methods to identify messages with adverse events mentions have been developed (eg, [7]).

Social media holds a lot of promise in improving communication and patient engagement [8]. Horvath et al [9] and Taggart et al [10] showed that information sharing and socializing with others were the criteria most often cited when HIV patients describe an ideal social network. Wang et al [11] modeled the discussions and interests of users of a forum for pregnant women using a topic model and showed that the women were sharing their experiences, fears, and concerns about medications. Stellefson et al [12] reviewed Web 2.0 interventions proposing a program of self-management to patients older than 50 years for their chronic disease. Patients highlighted the benefit of interacting with other patients. For example, sharing information through social networks enabled patients to communicate better with HCPs. Patients often use social media to discuss drug side effects and adherence to therapies. Mao et al [13] studied the messages from breast cancer patients treated by aromatase inhibitors. A total of 18.17% (4589/25,256) of the posts mentioned at least one adverse effect, and almost 12.8% (110/862) of the individuals mentioned discontinuing aromatase inhibitors. Chary et al [14] studied correlations between geographic distribution of prescription opioid misuse estimated from social media and the National Survey on Drug Usage and Health (NSDUH). They concluded that mentions of drug misuse on Twitter correlated strongly with the NSDUH estimates of opioid misuse.

Social media may even impact treatment adherence. In the study by Horvath et al [15], the results of a Web-based survey for HIV patients showed that 52.6% (164/312) of the participants were considered noncompliant. The meta-analysis published by Taggart et al [10] identified 2 studies on HIV populations that demonstrated a link between the use of social media and the improvement of compliance to treatment among users. Moreover, Mao et al [13] showed that breast cancer patients offer practical strategies to deal with drug side effects and provide support to each other. For example, 28.10% (7097/25,256) of the posts mentioned some method for addressing their aromatase inhibitor-related arthralgia, including exercising and pharmaceuticals, whether prescribed or over the counter.

Analysis of a huge number of narratives requires automated text mining techniques [5]. These techniques have been used to extract information from electronic health records. For example, Topaz et al [16] mined clinical narratives to identify heart failure patients who did not comply with their treatment. As for health records, detection of nonadherence behaviors in social media also requires text mining techniques.

Topic models could be used to discover hidden semantic structures in large sets of messages from social media. They could provide deeper exploration of nonadherence behaviors. This exploration is based on patient testimonies of their own decisions about drugs in real life.

Objective

Our objective was to evaluate a topic model approach to identify messages describing noncompliant behaviors regarding medications. Topics correspond to clusters of words that represent the themes addressed by the patients. The distributions of these themes in a corpus of messages are expected to enable the targeted extraction of posts corresponding to noncompliance behaviors. We focused on two noncompliant behaviors: (1) dose change and (2) treatment cessation.

Prior Work

Topic modeling is a text mining method designed for exploring the main topics that occur in a set of documents. With topic models, words that often occur together in text are grouped into different *topics*. On the basis of these topics, topic models provide a tool for unsupervised classification of massive collections of documents. Latent Dirichlet allocation (LDA) was developed by Blei et al as "...a generative probabilistic model for collections of discrete data such as text corpora..." [17].

Topic modeling algorithms have been used to analyze the thematic composition of text corpora extracted from social media in a variety of domains such as politics [18]. Several

authors explored tweets content using LDA to identify health topics, including tobacco use [19], seasonal influenza and allergies [15], and childhood obesity [20]. Sullivan et al analyzed users' comments from amazon to build a scoring system for food supplements [21].

Patient forums have been also explored using LDA. Yang et al [7] analyzed 1500 messages from patient forums to detect adverse drug reactions. The distributions of the themes obtained by applying the LDA model to this corpus made it possible to use similarity measurements for the annotated corpus compared with new messages. The authors proposed a message classifier based on these measurements. Noticeably, all the studies described above used messages in English.

With the objective of analyzing patients' QoL in breast cancer, Tapi Nzali et al [22] investigated posts from Facebook groups and a public French breast cancer forum using LDA modeling. They analyzed messages in French.

Several algorithms may be applied to use topic models. The original version of LDA modeling proposed by Blei et al [17] has been widely used (eg, [7,19,20,22,23]). Paul and Dredze developed extensions of the LDA model [15,24,25]. To establish their *Ailment Topic Aspect Model* (ATAM), they added several components to associate a term with a theme (eg, a disease), or consider it as not relevant. Then, based on 144 million tweets, they estimated general themes and disease-specific themes such as influenza, cancer, and dental problems. The semantic coherence of the topics obtained by ATAM was better for 61% (11/18) of the estimated topics compared with the LDA [25].

In this study, we investigate the use of LDA to analyze the themes in patient posts and identify noncompliance cases. To the best of our knowledge, this is the first study aiming at identifying forum posts related to nonadherence behaviors.

Methods

A summary of the approach presented in this study is provided in Figure 1.

Materials

The data was extracted from the Detec't database [26], a database developed by Kappa Santé [27] that collects messages from several French forums using a Web crawler. Detec't extracts messages from forums based on a named entity recognition module using a drug lexicon made by Kappa Santé and a fuzzy matching algorithm. The lexicon was based on Racine Pharma and the Anatomical Therapeutic Chemical (ATC) classification system [28]. Racine Pharma is an extensive source of drug names that covers all medications available on the French market, including brand names and active ingredients. Racine Pharma entries are mapped to the ATC.

More precisely, we extracted two corpora from Detec't: the first one corresponding to the messages related to escitalopram, an antidepressant drug, the other one related to aripiprazole, an antipsychotic drug. Rationale for choosing these drugs is that nonadherence cases are more likely to be found in chronic diseases and is a major concern in psychiatric disease management [29]. Moreover, these drugs belong to two different therapeutic classes: escitalopram is in a class of antidepressants called selective serotonin reuptake inhibitors; aripiprazole belongs to the so-called *atypical* second generation antipsychotics and acts as a partial dopamine agonist.

All messages extracted from Detec't database in this study were posted from 2004 to 2013 on three of the most popular French forums (doctissimo, atoute, and santé médecine). The metadata accompanying each message that form the corpus were as follows: (1) an identifier, (2) the date of publication on the forum, and (3) the forum from which the message was extracted. Messages were extracted based on the respective brand names: Seroplex and Abilify of the drugs. Posts were selected based on the presence of the drug name in the message.

Methods Used

Preliminary Data Processing

Preprocessing

The aim of the preprocessing step is the data cleaning to reduce noise and incoherence [30]. Preprocessing was done in six steps:

1. Considering that the R software (The R Project for Statistical Computing, Vienna) discriminates between lowercase and uppercase words, all messages were converted to lower case text.
2. The punctuation and stop words were removed.
3. We removed all instances of the drug name that was used to build the corpus (eg, seroplex). As it was present in each message, it was overrepresented and does not carry any further information.
4. Spaces were removed whenever needed to create *tokens*.
5. The stemming of words was carried out using Porter's algorithm [31,32].
6. We decided to keep *unigrams* and *bigrams*. This made it possible to retain frequent contiguous sequences of two items, such as *effets secondaires* (AEs).

Standardization of Dosage Mentions

As variations in representing dosage in posts are possible (eg, milligram or mg), we replaced it by a standard expression in the messages: we identified dosage mentions (eg, *10 mg*) by searching each sequence of numbers followed by a dosage unit. Then, we replaced the dosage mention by a neutral string of characters *dosemilligrams*.

Model Estimation

Document-Term Matrix Weighting

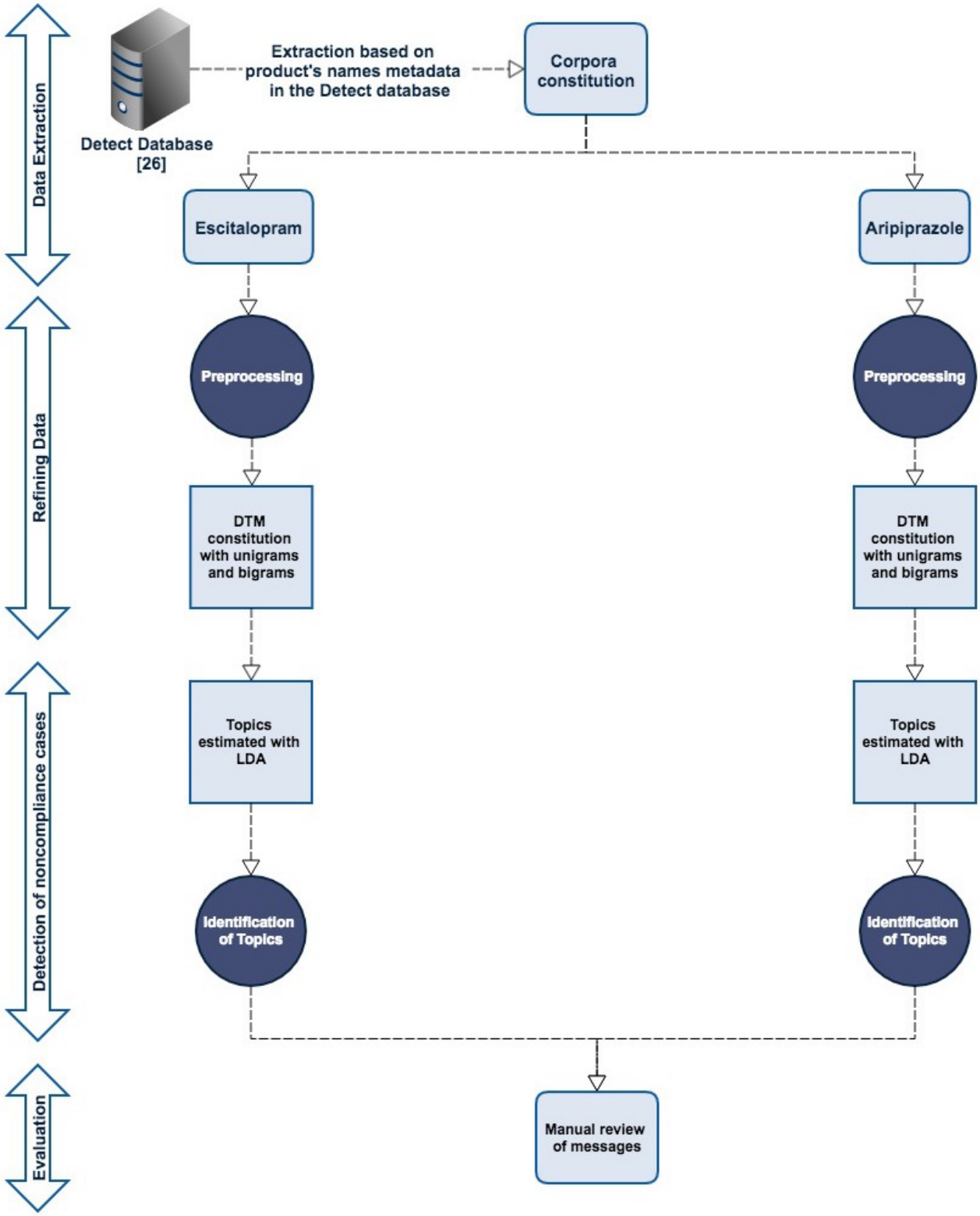
The *document-term matrix* (DTM) describes the frequency of terms that occur in the collection of posts: rows correspond to posts (documents), and columns correspond to terms. If a term occurs in a particular post, then the matrix entry corresponding to that row and column is 1, if not it is 0. The sparsity corresponds to the frequency of zero-valued elements in the matrix.

A maximum sparsity threshold, above which the token was removed, was determined empirically. The total sparsity of the matrix was calculated for an interval of sparsity thresholds applied to the columns. These values ranged from 99.95% to

80% and decremented by 0.025%. We included tokens corresponding to a DTM sparsity of at least 97%. Then, to avoid overrepresentation of frequent tokens, we applied a weighting to our DTM based on the *term-frequency-inverse-document-frequency* approach [33]. One DTM was generated for each corpus (escitalopram and aripiprazole, respectively) and used as input of the topic modeling.

To remove the tokens that corresponded to spelling errors or abbreviations and consider only words frequently used by patients, we removed infrequent tokens based on DTM sparsity.

Figure 1. Summary diagram.



Latent Dirichlet Allocation Modeling

In this study, we decided to work with LDA algorithm. The model was described as follows by Blei and Lafferty [34]:

> *In LDA, the observed data are the words of each document and the hidden variables represent the latent topical structure, i.e., the topics themselves and how each document exhibits them...The interaction between the observed documents and hidden topic structure is manifest in the probabilistic generative process associated with LDA.*

A document is a mixture of topics; that is, it corresponds to a probability distribution over all topics in the corpus. In other terms, when a patient writes a message, she or he decides to talk about a certain number of topics. When she or he talks about a topic in a message, she or he takes words with a certain probability from the set of terms that correspond to that topic. Assuming that model, each message contains several topics among all the identified topics, and the probability distribution shows how prominent the identified topics are in this message.

From a technical standpoint, rationale for choosing LDA was threefold:

- Compared with other types of topic modeling (latent semantic analysis, LSA; nonnegative matrix factorization, NMF; or singular value decomposition applied in the context of LSA), LDA methods are more suited in domains where data is in *semantic units*, such as words.
- LDA provides better interpretability of topics than other types of topic modeling (such as NMF).
- LDA also provides a better semantic coherence of estimated topics than LSA [35].

More precisely, we applied topic modeling with LDA algorithm developed by Blei et al [17,34]. The LDA model was estimated using the maximum a posteriori (MAP) algorithm described by Taddy [36,37]. MAP algorithm is a variant of expectation-maximization (EM) algorithm with a lower calculation cost and more stable results than the algorithms commonly used for estimates (Gibbs sampling, variational EM). At each iteration, instead of approximating the maximization of marginal likelihood, a combined estimate of the parameters is calculated by block-diagonalization of the Hessian matrix. This leads to an exact estimate of the distribution of topics, rather than an approximation. The number of topics was selected using the log Bayes factor [36]. Log Bayes factor is a ratio of likelihood used for model comparisons. By computing it against a one-topic model for several numbers of topics, it allows to select the most appropriate number. The output is twofold: (1) the probabilities of appearance associated with vocabulary terms in each of the *topics* and (2) the distribution thereof in the messages.

With the aim of optimizing interpretability and semantic coherence of topics, we considered a message significantly associated to a topic when at least 25% of the tokens it contained were associated to this topic. The 25% threshold was set empirically.

Evaluation

The aim of the evaluation step was to assess the number of messages correctly identified by our approach. Manual evaluation was performed in two steps:

1. We reviewed manually all messages related to the topics of interest (*dosage variation* and *treatment interruption*) in the two corpora (Escitalopram and Aripiprazole). A message is considered correctly classified if it describes a noncompliant behavior corresponding to the recognized topic. The evaluation of our classification was measured by the ratio of correctly classified messages for each topic of interest. Two annotators (RA and PF) participated to the review. To measure interannotator agreement (IAA), the two annotators evaluated a random selection of 20% of posts from each set of messages identified by the noncompliance topics. The IAA was calculated using Cohen kappa coefficient [38].
2. To estimate sensitivity or recall of our method, we randomly extracted 20% of the messages related to topics other than noncompliance (345/1723 messages for aripiprazole and 650/3246 for escitalopram). We manually classified them in two categories: messages with noncompliance behaviors and without.

Software

Analyses were performed using the R software. For the preprocessing of the corpus, the packages *tm* [39], *SnowballC*, and *slam* were used. Topic models were estimated using the following packages: *topicmodels* [40] and *MAPTPX*.

Results

Datasets Characteristics

Table 1 shows the number of messages in each corpus.

The preliminary preprocessing of escitalopram corpus returned a DTM of 3650 messages and 155,883 tokens (*unigrams* and *bigrams*). Setting the sparsity threshold at 99.35% (3626.275/3650), we obtained a DTM of 3649 messages and 1497 tokens. One message was removed because the terms it contained were particularly misspelled.

The processing of the aripiprazole corpus yielded a DTM of 2164 messages and 81,371 tokens. On the basis of a sparsity threshold of 99.25% (2147.77/2164), we obtained a DTM of 2164 messages and 1062 terms.

The *tokens* that appeared least frequently in the corpora were removed (Table 2).

Dosage Variations and Treatment Discontinuation

Model Estimation

The log Bayes factor topic selection method returned a total of 13 topics for the escitalopram corpus, as shown in Figure 2. The same approach led us to identify 11 topics for the aripiprazole corpus.

We obtained a total of 2691 messages evoking escitalopram and belonging to 13 topics. The 958 remaining messages were below the threshold regarding the association between terms

and topics, which was set at 25%. The average number of topics per message was 1.22 and the median 1.

For the aripiprazole data, we obtained a total of 1778 messages mentioning the drug and distributed among 11 topics. The 396 remaining messages were below the threshold for association between message terms and topics. The average number of topics per message was 1.31 and the median 1.

Topics Interpretation

As a topic must be interpretable with the first terms obtained (ranked by their probability of appearance) [34], topics found were labeled manually based on the first 15 words.

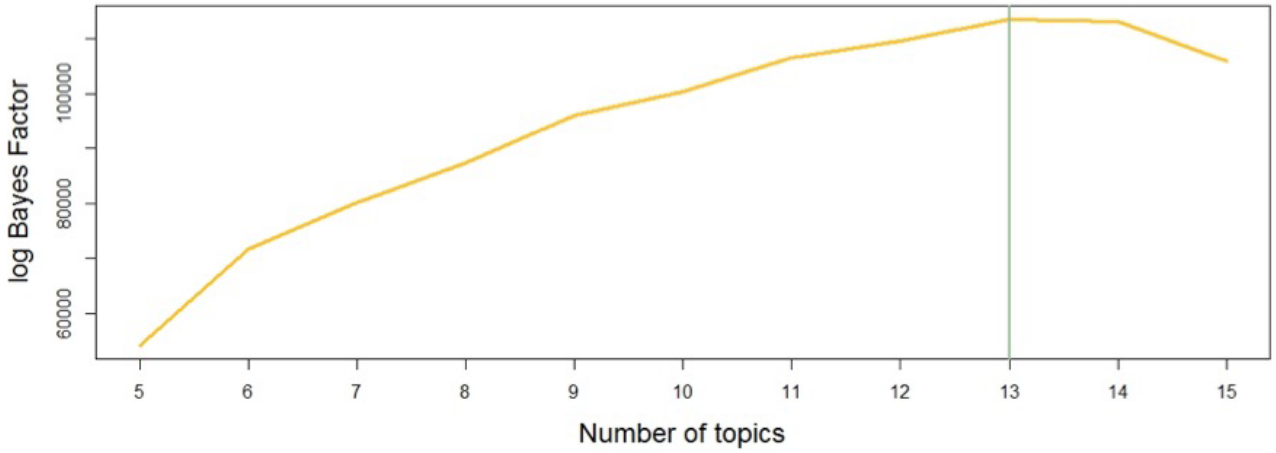
Table 1. Corpora description.

Drug	Therapeutic class	Number of messages containing the drug name, n	Date of publication
Escitalopram	Antidepressant	3650	2004 to 2013
Aripiprazole	Antipsychotic	2164	2005 to 2013

Table 2. Description of the document-term matrix (DTM) dissemination thresholds.

Drug	Term frequency before processing, n	Sparsity before processing, n (%)	Sparsity threshold per token, n (%)	Term frequency after processing, n	Sparsity after processing, n (%)
Escitalopram	155,883	155,774 (99.93)	3626.275 (99.35)	1497	151,097 (96.93)
Aripiprazole	81,371	81,281 (99.89)	2147.77 (99.25)	1062	78,922 (96.99)

Figure 2. Number of topics selected for Escitalopram using the log Bayes factor.



Escitalopram Topics

The list of topics and the distribution of messages in topics regarding escitalopram are displayed in Multimedia Appendix 1. This result is expressed as frequencies and proportions of messages (in relation to the 3649 escitalopram messages) associated with each theme. A message is associated with a topic if it contains at least 25% of terms for which the corresponding latent variability describes an association with the topic in question.

We noticed the emergence of a class containing messages describing user's experiences with the drug in a general way (topic 7) and how it affects their condition (topic 6). Topics 3 and 8 related to the day-to-day feeling of patients and the activities they have. Topic 2 was linked to the drug prescription by HCPs and topic 5 to panic attacks and anxiety. Topics 9, 10, and 12 focused on messages about AEs experienced or feared by users, along with the drug's effects overall. Topic 13 was related to the duration of the treatment.

Topics 1 was labeled as *general themes*. It describes themes associated with discussions between individuals and corresponds to poorly informative vocabulary. Such a collection of words provided no information of interest for our study. Nevertheless, it was widely used in messages, which explains the relatively high proportion of messages associated to this topic.

Messages about problems with treatment discontinuation and dosage variations were respectively included in topics 4 and 11. The intersection of the two noncompliance topics corresponded to 7 messages.

Aripiprazole Topics

The topics obtained by reproducing the modeling steps with the aripiprazole corpus are described in Multimedia Appendix 1.

Among the topics estimated for the aripiprazole corpus, we found the description of the patient's experience (topic 4) of his or her treatment. Three topics described the effects thereof (topics 3, 4, and 8), and one related to its duration (topic 9). Two topics focused on the patients' relationships with HCPs (topic 5) and other individuals (topic 6). Topic 7 described

treatment interruption. Dosage variations were described in topic 1. The intersection of the two noncompliance topics (7 and 1) corresponded to 6 messages. As for escitalopram, two topics were composed of noninformative words (general themes).

Multimedia Appendix 2 shows the identified topics in the two corpora.

Evaluation of the Approach

IAA rate was measured on 20% (169/844) of the messages identified by the noncompliance topics using Cohen kappa coefficient. We obtained a kappa of 0.90 (152/169).

Table 3. Annotations of the escitalopram corpus.

Identified behavior	Number of messages, n	Number of correctly classified messages, n	Precision, %
Dosage variations	187	54	28.9
Treatment cessation	216	100	46.3

Table 4. Annotations of the aripiprazole corpus.

Identified behavior	Number of messages, n	Number of correctly classified messages, n	Precision, %
Dosage variations	176	56	31.8
Treatment cessation	265	62	23.4

We calculated the ratio of messages corresponding to a case of noncompliance associated with each topic of interest. The results are displayed in Tables 3 and 4. Globally, the precision score for noncompliance was 32.6% (272/844). We obtained the lowest score (23.4%, 62/265) for the aripiprazole discontinuation topic and the highest score (46.3%, 100/216) for escitalopram discontinuation.

The analysis of 20% (345/1723 for aripiprazole, 650/3246 for escitalopram) of the messages related to other topics than noncompliance revealed only four messages describing a nonadherence behavior and not detected by our approach. The four false negative messages were all about stopping aripiprazole. Regarding the different subsets, we obtained a 94% (62/66) recall score for the aripiprazole cessation topic and 100% for the other ones. Globally, the estimated recall score was 98.5% (272/276).

We present below a detailed analysis of the results for the aripiprazole corpus.

Dosage Variations

Topics modeling identified 176 messages as *dosage variations* messages. Manual review revealed that only 56 (31.8%, 56/176) messages contained a true noncompliance declaration (2.6% of the 2164 posts initially in the corpus).

Among the 120 remaining messages, 68 (56.7%, 68/120) were discussions between patients comparing their dosages for aripiprazole. A total of 13 messages (10.8%, 13/120) contained information regarding the dosages of other prescribed drugs in addition to aripiprazole. The most cited drugs were amisulpride and olanzapine. Eight posts (6.7%, 8/120) were questions about aripiprazole's dosages, seven messages (5.8%, 7/120) evoked dosages modification, six posts (5.0%, 6/120) reported a dosage modification in agreement with the physician, and four messages (3.3%, 4/120) were advices.

Eight messages (6.7%, 8/120) did not contain dosage mentions but only variation words such as *increase* or *decrease*, usually accompanying dosage references.

The remaining six posts (5.0%, 6/120) mentioned variations that were planned or could occur in the future. For example:

> *[...] take 5 mg also for the moment the psy wants to increase the dose to 10 mg at the next appointment [...]*

In all the noncompliance cases (56 cases), the patient decreased the dose because of adverse drug reactions (ADRs). The most frequent ADRs mentioned in these posts were insomnia, asthenia, and libido problems.

Treatment Cessation

We identified 62 messages corresponding to noncompliance behaviors out of the 265 posts related to the *treatment cessation* topic (23.4%, 62/265). In other terms, 2.86% (62/2164) of the 2164 posts in the corpus are messages from patients taking aripiprazole who decided to stop their treatment.

Among the 203 remaining posts, thirteen posts (6.4%, 13/203) corresponded to the interruption of aripiprazole but were not cases of noncompliance: either aripiprazole was stopped to start another treatment, or the treatment cessation was decided by the physician.

A total of 89 posts (43.9%, 89/203) were written by patients who were prescribed this treatment in the past.

In 55 messages (27.1%, 55/203), the patient mentioned that she or he was reluctant to continue the treatment, mainly (74.6%, 151/203) because of ADRs. The most cited symptoms were insomnia, tiredness, libido problems, and nausea.

In 23 (11.3%, 23/203) posts, the patient was given more than one drug, and the post described the interruption of one of the other drugs (eg, in fourteen cases it was olanzapine that was stopped).

Eighteen posts were assigned erroneously to the treatment cessation topic because they contained terms like *stop*, although not reporting discontinuation of aripiprazole. These included 11 messages (5.4%, 11/203), where interruption was not related to any health topic such as in "[...] I'm stopped, like frozen,

[...],” five messages corresponding to cessation of alcohol, narcotics, or smoking (and not aripiprazole; 2.5%, 5/203), and two posts (<1%, 1/203) where the patient stopped her or his diet or other activities.

The five remaining messages (2.5%, 5/203) were requests for advices mentioning a possible treatment cessation.

Discussion

Principal Findings

Our study shows that topic models are useful to identify subsets of messages reporting noncompliance behaviors.

The topic models approach detected cases of noncompliance behaviors with averages recall and precision scores of 98.5% (272/276) and 32.6% (272/844), respectively. We concluded that the topic modeling presented in our study was a valuable sensitive method to detect noncompliance. However, it lacks specificity. We identified several situations leading to false positives: (1) two experiencers in the same message (eg, Peter takes 100 mg, whereas John takes 200 mg); (2) events in different time slots (eg, the patient reports that the doctor wants to increase or decrease the dose at the next appointment); and (3) the action concerns something else than the drug (eg, another medication and smoking cessation). Moreover, in several false positives, cessation or modification was prescribed by the physician.

Clinical Significance

We focused on escitalopram and aripiprazole used to treat depression and psychotic conditions, respectively.

Escitalopram

Almost one million individuals (2% of the overall population) initiated an antidepressant in France in 2011 [41]. Patients' adherence to antidepressant therapy must be evaluated. The reasons behind patient nonadherence to antidepressants include patient factors (eg, concerns about side effects and fears of addiction), as well as poor follow-up by the clinician and lack of sufficient patient education [42]. Better understanding of the patients' concerns about these medications can be achieved by exploring the messages in social media. We retrieved 2691 messages about escitalopram, among which 154 (5.71%, 154/2691) were noncompliance messages (Table 3). AEs were the most commonly cited reason for discontinuation and dose reduction. The more common side effects for escitalopram included nausea, weakness, dizziness, sleeping disorders, and sexual problems.

Aripiprazole

In a recently published review [43], a positive attitude toward medication at baseline in combination with good psychosocial function was the best predictor of objectively measured mean adherence over a 12-month period in patients with schizophrenia. AEs such as patient-reported cognitive impairment resulting from antipsychotic medication were predictors of nonadherence. Common side effects of aripiprazole also include weight gain, nausea, vomiting, changes in appetite, dizziness, drowsiness, feeling tired, and insomnia, among others. In our corpus, patients reported that such AEs were reasons for stopping the treatment or changing the dose.

Almost 7% (6.86%, 122/1778) of the posts in the aripiprazole corpus corresponded to noncompliance behaviors. All decisions to change the dose by the patient corresponded to decreasing the dose because of AEs. This result suggests that text mining methods must extract ADR information along with noncompliance annotation.

We calculated the rate of messages describing an effective noncompliance behavior. These rates were measured on messages corresponding to topics identified on the aripiprazole corpus. This evaluation resulted in 31.8% (56/176) for dosage variations and 23.4% (62/265) for treatment discontinuation.

Using topic models seems to be insufficient for identifying noncompliance cases on social media without a manual review step. However, this lexical approach produced only four false negatives and enabled us to reduce the corpus by focusing on messages that had a high probability to contain descriptions of targeted noncompliance behaviors.

Limitations

Our study focused on two drugs from two distinct classes. Both drugs are used to treat psychiatric disorders. A review currently including 50 clinical studies and 9476 participants taking antipsychotic drugs revealed an overall attrition from the included studies of 49% [44]. Consequently, our results regarding the noncompliance rate and the reasons for not being compliant cannot be extrapolated to other patient profiles. Further studies on other therapeutic classes must be conducted.

Manual review was required to distinguish between true and false positives in each dataset. The vocabulary used to describe dosage modifications or treatment interruptions in messages is commonly employed for characterizing other kinds of general variations or cessations (diet, smoking, etc). Topic models demonstrated their ability to identify potential noncompliance messages (average recall 98.5%, 272/276). Syntactic and semantic methods could be developed to recognize the experiencers, the temporal features, and the object concerned by the action in the sentences. Such methods could be applied to the datasets identified by the topic models to reduce the number of false positives and improve the precision score.

Another limitation of our work is the empirical determination of the thresholds used in our method. The thresholds concern the reduction in the size of the DTM and the significance of the association of messages to topics:

1. The choice of a sparsity threshold under 97% for DTM does not guarantee the best compromise between the calculation cost and the preservation of information for all the corpora we used.
2. The threshold for association between message words and topics, which was set at 25%, led to 23.23% (1354/5813) of messages not related to any topic.

Such empirical approach in the application of these methods is frequently reported in the literature; for example, Prier et al [19] set a suitable number of topics for their corpus by testing thresholds set every 50 topics.

Comparison With Other Work

Our study, to our knowledge, is the first one aiming at analyzing noncompliance behaviors from social media messages.

Most of the studies [18,21,22,25] used topic models to automatically label sets of tweets. Only 2 studies [7,22] focused on medical themes and messages from Web forums. Both used the same LDA model. Tapi Nzali et al [22] used the same R package [40]. However, their study design was different: they evaluated the correspondence between identified topics and QoL questionnaires, whereas our study aimed at detecting nonadherence behaviors.

Yang et al reported higher precision rates in their study [7]. Nevertheless, the aim of their study was detection of ADRs, not noncompliance practices.

Our approach could benefit from a more sophisticated model. The Structural Topic Model, developed by Wang et al [11], enables the modeling of correlations between topics and transitions made within messages. The additional components would enable the identification of relations between noncompliant practices and information, such as ADRs. We could therefore determine potential causes of nonadherence to treatment for each kind of drug.

Conclusions

Topic distributions in messages are a way to classify posts and detect noncompliance behaviors. The topic modeling approach achieved very high recall (98.5%, 272/276). Manual review of the messages in the noncompliance topics showed that almost 6.17% (276/4469) of the posts written by patients taking aripiprazole or escitalopram revealed noncompliance to treatment (half of them stopping their treatment). These findings indicate that social media mining may contribute to better understand noncompliance attitudes.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Exhaustive description of identified topics for each medication.

[XLSX File (Microsoft Excel File), 12KB - jmir_v20i3e85_app1.xlsx]

Multimedia Appendix 2

Topics found in each corpus ranked by the number of messages.

[XLSX File (Microsoft Excel File), 31KB - jmir_v20i3e85_app2.xlsx]

References

1. World Health Organization. Adherence to Long-term Therapies: Evidence for Action. Geneva: WHO; 2003.
2. Haynes RB, McDonald H, Garg AX, Montague P. Interventions for helping patients to follow prescriptions for medications. Cochrane Database Syst Rev 2002(2):CD000011. [doi: 10.1002/14651858.CD000011] [Medline: 12076376]
3. 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]
4. Fox S, Duggan M. PewInternet. 2013 Jan 15. Health Online 2013 URL: http://www.pewinternet.org/2013/01/15/health-online-2013/ [accessed 2018-02-27] [WebCite Cache ID 6xXiCgyLK]
5. Lardon J, Abdellaoui R, Bellet F, Asfari H, Souvignet J, Texier N, et al. Adverse drug reaction identification and extraction in social media: a scoping review. J Med Internet Res 2015 Jul 10;17(7):e171 [FREE Full text] [doi: 10.2196/jmir.4304] [Medline: 26163365]
6. Abou Taam M, Rossard C, Cantaloube L, Bouscaren N, Roche G, Pochard L, et al. Analysis of patients' narratives posted on social media websites on benfluorex's (Mediator®) withdrawal in France. J Clin Pharm Ther 2014 Feb;39(1):53-55. [doi: 10.1111/jcpt.12103] [Medline: 24304185]
7. Yang M, Kiang M, Shang W. Filtering big data from social media--Building an early warning system for adverse drug reactions. J Biomed Inform 2015 Apr;54:230-240 [FREE Full text] [doi: 10.1016/j.jbi.2015.01.011] [Medline: 25688695]
8. Househ M, Borycki E, Kushniruk A. Empowering patients through social media: the benefits and challenges. Health Informatics J 2014 Mar;20(1):50-58. [doi: 10.1177/1460458213476969] [Medline: 24550564]
9. Horvath KJ, Danilenko GP, Williams ML, Simoni J, Amico KR, Oakes JM, et al. Technology use and reasons to participate in social networking health websites among people living with HIV in the US. AIDS Behav 2012 May;16(4):900-910 [FREE Full text] [doi: 10.1007/s10461-012-0164-7] [Medline: 22350832]
10. 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 02;17(11):e248 [FREE Full text] [doi: 10.2196/jmir.4387] [Medline: 26525289]

11. Wang T, Huang Z, Gan C. On mining latent topics from healthcare chat logs. J Biomed Inform 2016 Jun;61:247-259 [FREE Full text] [doi: 10.1016/j.jbi.2016.04.008] [Medline: 27132766]
12. 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]
13. Mao JJ, Chung A, Benton A, Hill S, Ungar L, Leonard CE, et al. Online discussion of drug side effects and discontinuation among breast cancer survivors. Pharmacoepidemiol Drug Saf 2013 Mar;22(3):256-262 [FREE Full text] [doi: 10.1002/pds.3365] [Medline: 23322591]
14. Chary M, Genes N, Giraud-Carrier C, Hanson C, Nelson LS, Manini AF. Epidemiology from tweets: estimating misuse of prescription opioids in the USA from social media. J Med Toxicol 2017 Dec;13(4):278-286. [doi: 10.1007/s13181-017-0625-5] [Medline: 28831738]
15. Paul M, Dredze M. Semanticscholar. 2011. A model for mining public health topics from Twitter URL: https://pdfs.semanticscholar.org/41cb/ae26fe87307e6878e87b0a08056206a5c4c1.pdf [accessed 2018-02-21] [WebCite Cache ID 6xPFXCK8k]
16. Topaz M, Radhakrishnan K, Lei V, Zhou L. Mining clinicians' electronic documentation to identify heart failure patients with ineffective self-management: a pilot text-mining study. Stud Health Technol Inform 2016;225:856-857. [Medline: 27332377]
17. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res 2003;3:993-1022.
18. Yıldırım A, Üsküdarlı S, Özgür A. Identifying topics in microblogs using Wikipedia. PLoS One 2016 Mar 18;11(3):e0151885 [FREE Full text] [doi: 10.1371/journal.pone.0151885] [Medline: 26991442]
19. Prier KW, Smith MS, Giraud-Carrier C, Hanson CL. Identifying health-related topics on Twitter. In: Salerno J, Yang SJ, Nau D, Chai SK, editors. Social Computing, Behavioral-Cultural Modeling and Prediction. SBP 2011. Lecture Notes in Computer Science, vol 6589. Berlin, Heidelberg: Springer-Verlag; 2011:18-25.
20. Ghosh DD, Guha R. What are we 'tweeting' about obesity? Mapping tweets with Topic Modeling and Geographic Information System. Cartogr Geogr Inf Sci 2013;40(2):90-102 [FREE Full text] [doi: 10.1080/15230406.2013.776210] [Medline: 25126022]
21. Sullivan R, Sarker A, O'Connor K, Goodin A, Karlsrud M, Gonzalez G. Finding potentially unsafe nutritional supplements from user reviews with topic modeling. Pac Symp Biocomput 2016;21:528-539 [FREE Full text] [Medline: 26776215]
22. Tapi Nzali MD, Bringay S, Lavergne C, Mollevi C, Opitz T. What patients can tell us: topic analysis for social media on breast cancer. JMIR Med Inform 2017 Jul 31;5(3):e23 [FREE Full text] [doi: 10.2196/medinform.7779] [Medline: 28760725]
23. Yıldırım A, Üsküdarlı S, Özgür A. Identifying topics in microblogs using Wikipedia. PLoS One 2016;11(3):e0151885 [FREE Full text] [doi: 10.1371/journal.pone.0151885] [Medline: 26991442]
24. Paul MJ, Dredze M. You Are What You Tweet: Analyzing Twitter for Public Health. 2011 Presented at: Fifth International AAAI Conference on Weblogs and Social Media; July 17–21, 2011; Barcelona, Spain URL: http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2880
25. Paul MJ, Dredze M. Discovering health topics in social media using topic models. PLoS One 2014 Aug 01;9(8):e103408 [FREE Full text] [doi: 10.1371/journal.pone.0103408] [Medline: 25084530]
26. Abdellaoui R, Schück S, Texier N, Burgun A. Filtering entities to optimize identification of adverse drug reaction from social media: how can the number of words between entities in the messages help? JMIR Public Health Surveill 2017 Jun 22;3(2):e36 [FREE Full text] [doi: 10.2196/publichealth.6577] [Medline: 28642212]
27. Kappa Santé. URL: https://www.kappasante.com/ [accessed 2018-02-21] [WebCite Cache ID 6xPFzNeah]
28. Bousquet C, Dahamna B, Guillemin-Lanne S, Darmoni SJ, Faviez C, Huot C, et al. The adverse drug reactions from patient reports in social media project: five major challenges to overcome to operationalize analysis and efficiently support pharmacovigilance process. JMIR Res Protoc 2017 Sep 21;6(9):e179.
29. McHorney CA, Spain CV. Frequency of and reasons for medication non-fulfillment and non-persistence among American adults with chronic disease in 2008. Health Expect 2011 Sep;14(3):307-320 [FREE Full text] [doi: 10.1111/j.1369-7625.2010.00619.x] [Medline: 20860775]
30. Iavindrasana J, Cohen G, Depeursinge A, Müller H, Meyer R, Geissbuhler A. Clinical data mining: a review. Yearb Med Inform 2009:121-133. [Medline: 19855885]
31. Porter MF. Cs.toronto.edu. 1980. An algorithm for suffix stripping URL: https://www.cs.toronto.edu/~frank/csc2501/Readings/R2_Porter/Porter-1980.pdf [accessed 2018-02-21] [WebCite Cache ID 6xPGZzLyA]
32. Savoy J. Light Stemming Approaches for the French, Portuguese, German Hungarian Languages. In: Proceedings of the 2006 ACM symposium on Applied computing. USA: ACM; 2006 Presented at: SAC 2006; April 23 -27; Dijon, France p. 1031-1035.
33. Salton G, Mcgill MJ. Introduction to Modern Information Retrieval. New York: McGraw-Hill, Inc; 1986.
34. Blei DM, Lafferty JD. CS Columbia. 2009. Topic models URL: http://www.cs.columbia.edu/~blei/papers/BleiLafferty2009.pdf [accessed 2018-02-21] [WebCite Cache ID 6xPHPCaI4]

35. Stevens K, Kegelmeyer P, Andrzejewski D, Buttler D. Exploring topic coherence over many models and many topics. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. 2012 Presented at: EMNLP-CoNLL '12; July 12 - 14, 2012; Jeju Island, Korea p. 952-961.
36. Taddy MA. On Estimation and Selection for Topic Models. In: Proceedings of the 15th International Conference on Artificial Intelligence and Statistics. 2012 Presented at: AISTATS, 2012; 2012; La Palma, Canary Islands.
37. Sievert C, Shirley KE. LDAvis: A method for visualizing and interpreting topics. In: Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces. 2014 Presented at: Workshop on Interactive Language Learning, Visualization, and Interfaces; June 27, 2014; Baltimore, Maryland, USA p. 63-70.
38. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960 Apr 01;20(1):37-46. [doi: 10.1177/001316446002000104]
39. Feinerer I, Hornik K, Meyer D. Text mining infrastructure in R. J Stat Softw 2018 Mar 31;25(5):2008.
40. Grün B, Hornik K. Topicmodels: an R package for fitting topic models. J Stat Softw 2011 May 09;40(13):1-30.
41. Fagot JP, Cuerq A, Samson S, Fagot-Campagna A. Cohort of one million patients initiating antidepressant treatment in France: 12-month follow-up. Int J Clin Pract 2016 Sep;70(9):744-751. [doi: 10.1111/ijcp.12850] [Medline: 27484351]
42. Sansone RA, Sansone LA. Antidepressant adherence: are patients taking their medications? Innov Clin Neurosci 2012 May;9(5-6):41-46 [FREE Full text] [Medline: 22808448]
43. Velligan DI, Sajatovic M, Hatch A, Kramata P, Docherty JP. Why do psychiatric patients stop antipsychotic medication? A systematic review of reasons for nonadherence to medication in patients with serious mental illness. Patient Prefer Adherence 2017;11:449-468 [FREE Full text] [doi: 10.2147/PPA.S124658] [Medline: 28424542]
44. Komossa K, Rummel-Kluge C, Hunger H, Schmid F, Schwarz S, Duggan L, et al. Olanzapine versus other atypical antipsychotics for schizophrenia. Cochrane Database Syst Rev 2010 Mar 17(3):CD006654 [FREE Full text] [doi: 10.1002/14651858.CD006654.pub2] [Medline: 20238348]

Abbreviations

ADR: adverse drug reaction
AE: adverse effect
ATAM: Ailment Topic Aspect Model
ATC: Anatomical Therapeutic Chemical
DTM: document-term matrix
EM: expectation-maximization
HCP: health care professional
IAA: inter-annotator agreement
LDA: latent Dirichlet allocation
LSA: latent semantic analysis
MAP: maximum a posteriori
NMF: nonnegative matrix factorization
NSDUH: National Survey on Drug Usage and Health
QoL: quality of life
WHO: World Health Organization

Edited by G Eysenbach; submitted 22.10.17; peer-reviewed by E Chavarria, C Bousquet, A Keepanasseril; comments to author 16.11.17; revised version received 27.12.17; accepted 13.01.18; published 14.03.18.

Please cite as:
Abdellaoui R, Foulquié P, Texier N, Faviez C, Burgun A, Schück S
Detection of Cases of Noncompliance to Drug Treatment in Patient Forum Posts: Topic Model Approach
J Med Internet Res 2018;20(3):e85
URL: http://www.jmir.org/2018/3/e85/
doi:10.2196/jmir.9222
PMID:29540337

Original Paper

Prenatal Remote Monitoring of Women With Gestational Hypertensive Diseases: Cost Analysis

Dorien Lanssens[1,2], MSc; Thijs Vandenberk[1], MSc; Christophe JP Smeets[1,3], MSc; Hélène De Cannière[1,3], MSc; Sharona Vonck[2,4], PhD; Jade Claessens[1], BSc; Yenthel Heyrman[1], BSc; Dominique Vandijck[4,5], PhD; Valerie Storms[1], PhD; Inge M Thijs[1,6], PhD; Lars Grieten[1], PhD; Wilfried Gyselaers[1,2,7], PhD

[1]Mobile Health Unit, Faculty of Medicine and Life Sciences, Hasselt University, Hasselt, Belgium

[2]Department of Gynaecology, Ziekenhuis Oost-Limburg, Genk, Belgium

[3]Department of Cardiology, Ziekenhuis Oost-Limburg, Genk, Belgium

[4]Department of Health and Life Sciences, Hasselt University, Hasselt, Belgium

[5]Faculty of Medicine and Life Sciences, Ghent University, Gent, Belgium

[6]Future Health Department, Ziekenhuis Oost-Limburg, Genk, Belgium

[7]Department of Physiology, Hasselt University, Hasselt, Belgium

Corresponding Author:
Dorien Lanssens, MSc
Mobile Health Unit
Faculty of Medicine and Life Sciences
Hasselt University
Martelarenlaan 42
Hasselt, 3500
Belgium
Phone: 32 (0) 11 26 8 ext 32
Email: dorien.lanssens@uhasselt.be

Abstract

Background: Remote monitoring in obstetrics is relatively new; some studies have shown its effectiveness for both mother and child. However, few studies have evaluated the economic impact compared to conventional care, and no cost analysis of a remote monitoring prenatal follow-up program for women diagnosed with gestational hypertensive diseases (GHD) has been published.

Objective: The aim of this study was to assess the costs of remote monitoring versus conventional care relative to reported benefits.

Methods: Patient data from the Pregnancy Remote Monitoring (PREMOM) study were used. Health care costs were calculated from patient-specific hospital bills of Ziekenhuis Oost-Limburg (Genk, Belgium) in 2015. Cost comparison was made from three perspectives: the Belgian national health care system (HCS), the National Institution for Insurance of Disease and Disability (RIZIV), and costs for individual patients. The calculations were made for four major domains: prenatal follow-up, prenatal admission to the hospital, maternal and neonatal care at and after delivery, and total amount of costs. A simulation exercise was made in which it was calculated how much could be demanded of RIZIV for funding the remote monitoring service.

Results: A total of 140 pregnancies were included, of which 43 received remote monitoring (30.7%) and 97 received conventional care (69.2%). From the three perspectives, there were no differences in costs for prenatal follow-up. Compared to conventional care, remote monitoring patients had 34.51% less HCS and 41.72% less RIZIV costs for laboratory test results (HCS: mean €0.00 [SD €55.34] vs mean €38.28 [SD € 44.08], *P*<.001; RIZIV: mean €21.09 [SD €27.94] vs mean €36.19 [SD €41.36], *P*<.001) and a reduction of 47.16% in HCS and 48.19% in RIZIV costs for neonatal care (HCS: mean €989.66 [SD €3020.22] vs mean €1872.92 [SD €5058.31], *P*<.001; RIZIV: mean €872.97 [SD €2761.64] vs mean €1684.86 [SD €4702.20], *P*<.001). HCS costs for medication were 1.92% lower in remote monitoring than conventional care (mean €209.22 [SD €213.32] vs mean €231.32 [SD 67.09], *P*=.02), but were 0.69% higher for RIZIV (mean €122.60 [SD €92.02] vs mean €121.78 [SD €20.77], *P*<.001). Overall HCS costs for remote monitoring were mean €4233.31 (SD €3463.31) per person and mean €4973.69 (SD €5219.00) per person for conventional care (*P*=.82), a reduction of €740.38 (14.89%) per person, with savings mainly for RIZIV of €848.97 per person (23.18%; mean €2797.42 [SD €2905.18] vs mean €3646.39 [SD €4878.47], *P*=.19). When an additional fee of €525.07 per month

per pregnant woman for funding remote monitoring costs is demanded, remote monitoring is acceptable in their costs for HCS, RIZIV, and individual patients.

Conclusions: In the current organization of Belgian health care, a remote monitoring prenatal follow-up of women with GHD is cost saving for the global health care system, mainly via savings for the insurance institution RIZIV.

(J Med Internet Res 2018;20(3):e102) doi:10.2196/jmir.9552

KEYWORDS

remote monitoring; gestational hypertensive diseases; reimbursement; cost-effectiveness

Introduction

Remote monitoring in obstetrics is a relatively new field of research; only a few trials have shown the effectiveness of remote monitoring in obstetrical care for both mother and child. When uterine activity is transmitted by telecommunication, significant prolonged pregnancy survivals are observed [1,2]. Higher feelings of self-efficacy and a reduction in (unscheduled) face-to-face visits [3-6] is reported when remote monitoring is used in the prenatal follow-up of pregnant women with gestational diabetes mellitus in comparison to conventional care. In addition, elevated feelings of maternal satisfaction were obtained when remote monitoring was used in obstetrical care [3,6-8]. Finally, the newborns did have a higher gestational age at delivery [9] and were less likely to be of low birth weight [1,9] or to be admitted to the neonatal intensive care unit [1,9,10] when a remote monitoring group was compared to a conventional care group. In an earlier publication, we reported that remote monitoring in pregnant women with gestational hypertensive diseases (GHD) reduces the number of inductions and maternal prenatal admissions [10]. However, until now, few studies have evaluated the economic impact of remote monitoring compared to conventional care [9,11], and no study is known about the cost-effectiveness of a remote monitoring prenatal follow-up program for women diagnosed with GHD.

The Pregnancy Remote Monitoring (PREMOM) study was designed for women diagnosed with GHD who had their prenatal follow-up in Ziekenhuis Oost-Limburg (Genk, Belgium). According to the Flanders' register of perinatal outcomes, the prevalence of hypertensive disorders in pregnancy is 4.6%: 0.3% deliver before 34 weeks, 0.6% deliver between 34 and 37 weeks, and 3.7% deliver after 37 weeks [12]. As a continuation of this trial, a study was designed with the objective of quantifying the costs of both remote monitoring and conventional care from the perspectives of the Belgium global health care system (HCS), which combines costs for the National Institution for Insurance of Disease and Disability (*Rijksinstituut voor Ziekte- en Invaliditeitsverzekering*; RIZIV) and costs for individual patients [13]. The calculations were made for four major domains: prenatal follow-up, prenatal admission to the hospital, maternal and neonatal care at and after delivery, and total amount of costs. A simulation exercise was made when an additional fee of €100 per month per patient for remote monitoring was charged. We hypothesized the addition of remote monitoring to a prenatal follow-up program for pregnant women with GHD to be cost-effective when compared to conventional care. This paper reports on the results for the Belgium situation.

Methods

Data

Data collected from the PREMOM study was used for this cost analysis. The PREMOM study design and data collection method are described in detail elsewhere [10]. Briefly, the PREMOM study was a 1-year retrospective study, performed in the outpatient clinic of a second-level prenatal center where pregnant women with GHD received remote monitoring or conventional care. From January 1 to December 31, 2015, 166 pregnant women were diagnosed with GHD: 53 of them received remote monitoring and 113 received conventional care. After excluding five patients in the remote monitoring group and 15 in the conventional care group because of missing data, 48 patients in the remote monitoring group and 98 in the conventional care group were included in the final analysis.

Women consenting for remote monitoring received obstetric surveillance using a Withings Wireless Blood Pressure Monitor, Withings Smart Body Analyzer, and a Withings Pulse O_2 (Withings, Issy-les-Moulineux, France). Pregnant women participating in the prenatal remote follow-up program were asked to perform one blood pressure measurement in the morning and one in the evening, one weight measurement a day, and to wear an activity tracker day and night until delivery or hospital admission. The data from the monitor devices were transmitted to a Web-based dashboard developed by the mobile health unit of Hasselt University. Predetermined alarm signals were set and alarm events were communicated with the obstetrician in charge to discuss management options before contacting and instructing patients at home. Therapeutic interventions were according to local management. The clinical goal of routine prenatal outpatient care is to timely detect an abnormal course of maternal and/or fetal health. The study protocol was approved by the local ethics committees responsible for the site. The investigation conformed to the principles outlined in the Declaration of Helsinki. All patients gave written informed consent, and data were treated confidentially.

Study Design

The objective of the study was to quantify the costs of remote monitoring versus conventional care from the perspectives of the HCS, the RIZIV, and the patients. The costs of the HCS are the total amount of costs that have to be paid to cover the care that has been provided. These HCS costs can be divided into two subgroups who have to pay their part of the costs: (1) RIZIV, the national institutional social security in Belgium, which ensures every insured individual, regardless of financial

situation, has access to necessary qualitative medical care in accordance with the tariff agreements between caregivers and government [14] and (2) the patients who have to pay their part of care from their own financial resources. The HCS costs are estimated by using the national tariffs applied for these services. The costs for the RIZIV were calculated using the Belgium national reimbursement tariffs [12]. The costs for the patients were the HCS costs minus the RIZIV costs. The four major domains in which the costs are divided and their subcategories are presented subsequently. A detailed overview of the included costs are presented in Multimedia Appendix 1.

Cost Analysis

Prenatal Follow-Up

All costs related to urgent and nonurgent in-office visits were used in the prenatal follow-up cost analysis: (1) costs of prenatal consultations, (2) costs of ultrasounds, and (3) costs of cardiotocographics.

Prenatal Admission to the Hospital

To evaluate the economic impact of remote monitoring on the three major stakeholders, the following data points were collected when the pregnant women were admitted to the prenatal ward: (1) costs related to the laboratory test results of the mother, (2) costs of the medicines, and (3) costs related to the admission.

Maternal and Neonatal Care at and After Delivery

For both groups, the following costs were included for this topic: (1) costs of the delivery, (2) costs necessary for the care of the neonate, and (3) other costs.

Total Amount of Costs

After analyzing the previously mentioned data, a cost analysis of the total amount of costs was made. This included (1) costs of the prenatal follow-up, (2) costs of the prenatal admission to the prenatal ward, and (3) costs of the maternal and neonatal care at and after delivery.

Simulation Exercise

A simulation exercise was made in which the amount that could be demanded by RIZIV for funding of the remote monitoring service was calculated. This charge was calculated by dividing the cost savings in RIZIV (by subtracting the total costs of the remote monitoring group from those of the conventional care group) by the mean time of prenatal remote monitoring follow-up per pregnant woman. This charge could be used to finance the costs which were needed to perform remote monitoring in the prenatal follow-up of women at risk for GHD, such as the need of midwives to accompany the pregnant women to their remote monitoring follow-up and to interpret the (alarm) signals, the need of obstetrics to refer and supervise the pregnant women at risk, and the need of technical staff to maintain the platform, to give technical support, etc.

Statistical Analysis

The baseline characteristics are continuous data summarized as mean and standard deviation. Categorical data are summarized as count and percentage and were compared using the chi-square test or Fisher exact test, when appropriate. Costs were reported as means, standard deviations, medians, and interquartile ranges. Cost data are typically highly skewed [14] because a few patients incur particularly high costs; therefore, the Mann-Whitney *U* test was used to compare costs across groups. Both univariate and multivariate analyses were performed for analyzing the costs for the three domains.

The nominal level alpha<.05 was considered significant. All statistical analyses were performed with SPSS release 24.0.

Results

Baseline Characteristics

The baseline characteristics of the patients are summarized in Table 1. Of the 48 patients participating in the remote monitoring study, five (10%) were excluded due to missing data. In the conventional care group, one participant was excluded due to missing data (1/98, 1%). Finally, the remote monitoring group consisted of 43 (30.7%) patients and the conventional care group had 97 (69.3%). The baseline clinical characteristics of the population enrolled were almost homogeneous, without differences between the two groups except for primigravida (44%, 19/43) in the remote monitoring group versus 66% (65/97) in the conventional care group (*P*=.02) and smoking (0%, 0/43) in the remote monitoring group versus 10%, (10/97) in the conventional care group (*P*=.03).

Table 1. Baseline clinical characteristics (N=140).

Variables	Remote monitoring group (n=43)	Conventional care group (n=97)	*P* value (2-tailed)
Age (years), mean (SD)	31.72 (4.44)	31.95 (4.77)	.77
Prepregnancy weight (kg), mean (SD)	70.12 (16.26)	76.80 (19.75)	.05
Height (cm), mean (SD)	165.65 (6.89)	167.08 (6.86)	.18
BMI (kg/m^2), mean (SD)	25.23 (5.03)	27.01 (6.94)	.32
Primigravida, n (%)	19 (44)	65 (66)	.02
Cardiovascular disorders, n (%)	0 (0)	1 (1)	.99
Coagulation disorders, n (%)	1 (2)	1 (1)	.52
Endocrine disorders, n (%)	2 (5)	5 (5)	.99
Immunology disorders, n (%)	1 (2)	2 (2.04)	.99
Smoker, n (%)	0 (0)	10 (10)	.03

Health Care Costs

The health care costs are presented in Table 2. The results are discussed in detail subsequently.

To investigate the influence of the maternal demographics and characteristics on the health care costs, a multiple linear regression analysis and a multivariate logistic regression analysis was performed. A detailed overview of these data are provided in Multimedia Appendix 2. No important influences of the maternal demographics and characteristics was found in the health care costs.

Cost Analysis

Prenatal Follow-Up

No differences were found in costs for prenatal follow-up (prenatal visits, ultrasounds, and costs of cardiotocographics): not in the costs for the HCS, the RIZIV, or the patients.

Prenatal Admission to the Hospital

Patients admitted to the remote monitoring group did have 34.51% less HCS and 41.72% less RIZIV costs for laboratory test results compared to conventional care group (HCS: remote monitoring mean €25.07, SD €55.34 vs conventional care mean €38.28, SD €44.08, *P*<.001; RIZIV: remote monitoring mean €21.09, SD €27.94 vs conventional care mean €36.19, SD €41.36, *P*<.001). Also, the HCS cost for the medicaments were 1.92% lower in the remote monitoring group compared to the conventional care group (mean €209.22, SD €141.86 vs mean €213.32, SD €67.09, *P*=.02), but the RIZIV costs were 0.69% higher in the remote monitoring group compared to the conventional care group (mean €122.60, SD €92.02 vs mean €121.76, SD €20.77, *P*<.001).

Maternal and Neonatal Care at and After Delivery

No differences were found in costs for delivery in the remote monitoring group versus the conventional care group. A reduction of 47.16% in HCS cost and 48.19% in RIZIV costs for neonatal care was found in the remote monitoring group compared to the conventional care group (HCS: remote monitoring mean €989.66, SD €3020.22 vs conventional care mean €1872.92, SD €5058.31, *P*<.001; RIZIV: remote monitoring mean €872.97, SD €2761.64 vs conventional care mean €1684.86, SD €4702.20, *P*<.001). Other costs were for the HCS 57.86% and RIZIV 58.63% lower in remote monitoring versus conventional care (HCS: remote monitoring mean €26.63, SD €11.83 vs conventional care mean €63.19, SD €158.23, *P*=.04; RIZIV remote monitoring mean €26.14, SD €19.86 vs conventional care mean €63.19, SD €158.23, *P*<.001), but 0.77% higher for the patients in remote monitoring versus conventional care (mean €0.49, SD €20.99 vs mean €0.00, SD €0.00, *P*=.01).

Total Amount of Costs

An overview of the total amount of costs is presented in Figure 1 and in Multimedia Appendix 3. There were no significant differences between remote monitoring and conventional care in total amount of costs for HCS (remote monitoring mean €4233.31, SD €3463.31 vs conventional care mean €4973.69, SD €5219.00, *P*=.82), the RIZIV (remote monitoring mean €2797.42, SD €2905.18 vs conventional care mean €3646.40, SD 4878.47, *P*=.19), or the patients (remote monitoring mean €1435.89, SD €829.09 vs conventional care mean €1327.30, SD €753.94, *P*=.38). But, a cost reduction of €740.38 per person (14.89%) was made for HCS and a cost reduction of €848.97 (23.18%) was made for RIZIV in remote monitoring compared to conventional care. Patient's costs were slightly higher (€108.59, 8.18%) for remote monitoring than for conventional care.

Simulation Exercise

A simulation exercise was made in which it was calculated how much could be demanded of RIZIV for funding the remote monitoring service. For this study, 43 pregnant women were included in the analysis with a range of 1 day of participation to 145 days of participation in the PREMOM project. The mean time of participation in this project was 44.42 days or 1.41 months (Multimedia Appendix 4). By dividing €740.35 by 1.41 months, a funding of €525.07 per month per pregnant woman could be asked. Because of the difference of almost €1000 per person in costs for the RIZIV, it was reasonable to charge the supplementary costs to RIZIV. As a result, there was a significant difference in costs for HCS of a reduction of €2.11 per person in remote monitoring versus conventional care (remote monitoring: mean €4971.58, SD 3479.69; conventional care: mean 4973.69, SD 5219.00, *P*=.01) and in RIZIV costs by also having a reduction of €110.70 per person in remote monitoring versus conventional care (remote monitoring: mean 3535.69, SD 2931.90; conventional care: mean 3646.39, SD 4878.47, *P*=.005). The patient still does not have to pay more for their prenatal care (remote monitoring: mean €1435.89, SD €829.09; conventional care: mean €1327.30, SD €753.94, *P*=.38). An overview of the costs is shown in Multimedia Appendix 4 and in Figure 2.

Table 2. Health care costs. All costs in euros. HCS: health care system; IQR: interquartile range; RIZIV: National Institution for Insurance of Disease and Disability.

Cost variable	Study group				Cost savings in Euros in remote monitoring group, n (%)	*P* value (2-tailed)
	Remote monitoring (n=43)		Conventional care (n=97)			
	Mean (SD)	Median (IQR)	Mean (SD)	Median (IQR)		
Prenatal follow-up						
Prenatal visits						
HCS	184.26 (79.10)	205.80 (144.06-226.38)	183.31 (71.79)	185.22 (144.06-226.38)	–0.95 (–0.52)	.71
RIZIV	110.58 (47.83)	123.50 (86.45-135.85)	110.00 (43.08)	111.15 (86.45-135.85)	–0.58 (–0.52)	.71
Patients	73.69 (31.87)	82.30 (57.61-90.53)	73.31 (28.71)	74.07 (57.61-90.53)	–0.38 (–0.52)	.71
Ultrasounds						
HCS	89.66 (58.61)	79.77 (79.77-106.36)	96.49 (57.23)	79.77 (79.77-106.36)	6.83 (7.08)	.96
RIZIV	81.30 (53.14)	72.33 (72.33-96.44)	87.49 (51.89)	72.33 (72.33-96.44)	6.19 (7.08)	.96
Patients	8.36 (5.47)	7.44 (7.44-9.92)	9.00 (5.34)	7.44 (7.44-9.92)	0.64 (7.08)	.96
Cardiotocographics						
HCS	127.58 (130.45)	124.68 (0.00-187.02)	93.19 (105.37)	62.34 (0.00-124.68)	–34.39 (–36.90)	.15
RIZIV	63.79 (65.22)	62.34 (0.00-93.1)	46.59 (52.68)	31.17 (0.00-62.34)	–17.20 (–36.90)	.15
Patients	63.79 (65.22)	62.34 (0.00-93.51)	46.59 (52.68)	31.17 (31.17-62.34)	–17.20 (-36.90)	.15
Prenatal admission						
Laboratory test results						
HCS	25.07 (55.34)	0.00 (0.00-19.58)	38.28 (44.08)	27.86 (5.13-56.74)	13.21 (34.51)	<.001
RIZIV	21.09 (27.94)	0.00 (0.00-19.07)	36.19 (41.36)	25.74 (5.13-50.53)	15.10 (41.72)	<.001
Patients	3.98 (14.06)	0.00 (0.00-0.00)	2.09 (8.78)	0.00 (0.00-0.00)	–1.89 (–90.43)	.78
Prenatal admission						
HCS	1423.57 (1184.78)	1166.62 (1013.25-1407.54)	1336.40 (670.99)	1172.61 (950.68-1450.04)	–87.17 (–6.52)	.73
RIZIV	798.47 (596.93)	663.30 (600.25-786.59)	783.44 (372.81)	714.96 (501.09-922.33)	–15.03 (–1.92)	.63
Patients	625.10 (606.57)	497.67 (394.29-617.61)	552.96 (372.50)	477.88 (324.57-663.41)	–72.14 (–13.05)	.41
Medicaments						
HCS	209.22 (141.86)	168.73 (155.71-206.18)	213.32 (67.09)	204.65 (168.99-233.79)	4.10 (1.92)	.02
RIZIV	122.60 (92.02)	106.03 (99.61-111.77)	121.76 (20.77)	114.81 (108.02-130.01)	–0.84 (–0.69)	<.001
Patients	86.61 (68.81)	63.71 (47.69-97.87)	91.56 (20.77)	79.13 (55.67-108.43)	4.95 (5.41)	.14
Maternal and neonatal care						
Delivery						
HCS	1157.66 (469.34)	1298.10 (670.34-1329.38)	1076.61 (485.14)	998.94 (670.34-1298.10)	–81.05 (–7.53)	.15
RIZIV	700.48 (186.41)	670.34 (370.34-685.98)	712.87 (196.03)	670.34 (663.34-755.66)	12.39 (1.74)	.79
Patients	457.17 (344.53)	627.76 (0.00-643.40)	363.73 (404.17)	424.11 (0.00-628.86)	–93.44 (–25.69)	.15
Neonatal care						
HCS	989.66 (3020.22)	146.32 (102.67-374.19)	1872.92 (5058.31)	290.78 (147.69-625.23)	883.26 (47.16)	<.001
RIZIV	872.97 (2761.64)	98.48 (85.49-279.14)	1684.86 (4702.20)	230.45 (104.81-519.38)	811.89 (48.19)	<.001
Patients	116.69 (263.74)	48.22 (13.01-95.05)	188.06 (413.95)	61.68 (23.69-120.19)	71.37 (37.95)	.10
Other						
HCS	26.63 (11.83)	25.73 (25.73-25.73)	63.19 (158.23)	25.73 (25.73-25.73)	36.56 (57.86)	.04
RIZIV	26.14 (19.86)	25.73 (21.10-25.73)	63.19 (158.23)	25.73 (25.73-25.73)	37.05 (58.63)	<.001
Patients	0.49 (20.99)	0.00 (0.00-0.00)	0.00 (0.00)	25.73 (25.73-25.73)	–0.49 (–0.77)	.01

Figure 1. Total amount of costs for remote monitoring (RM) and conventional care (CC) groups paid by health care service (HCS), National Institution for Insurance of Disease and Disability (RIZIV), and patients.

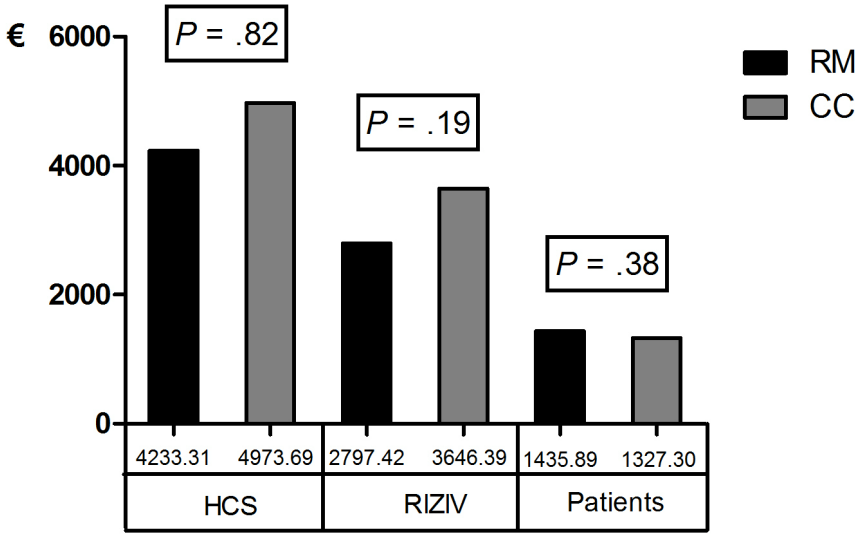
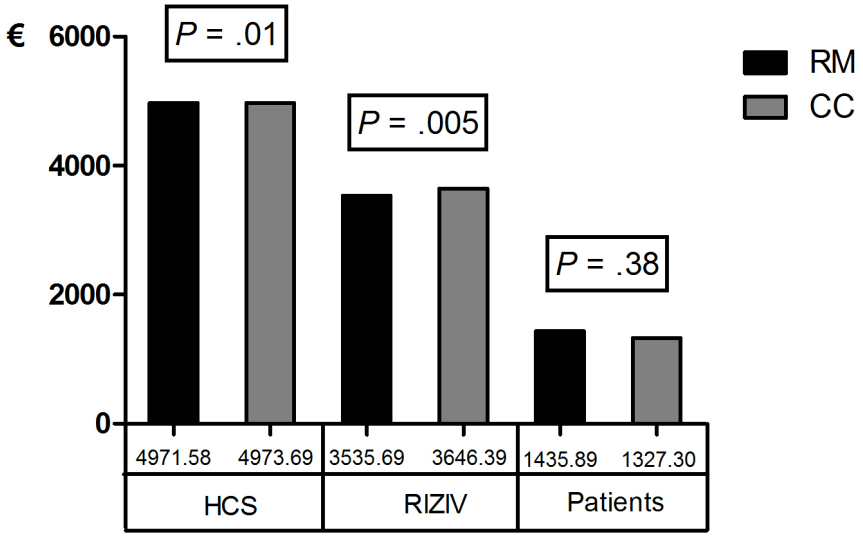


Figure 2. Total amount of costs plus remote monitoring for remote monitoring (RM) and conventional care (CC) groups paid by health care service (HCS), National Institution for Insurance of Disease and Disability (RIZIV), and patients.



Discussion

Principal Findings

The main finding of this study is that a remote monitoring prenatal follow-up for pregnant women at risk for GHD reduces the total amount of costs for national health care in comparison to a standard follow-up strategy. This cost reduction is due to a marked reduction in the consumption of health care services, including laboratory test results taken, medication use, and maternal and neonatal admissions. When an additional fee of €525.07 per month per pregnant woman for funding remote monitoring costs is asked, remote monitoring is still acceptable in their costs for HCS, RIZIV, and individual patients.

Strengths and Limitations

The use of “real-life” data from the hospital bills is the main strength of this study. By using these data, the actual situation of pregnancies complicated with GHD is simulated and these results are generalizable for settings with similar economics and social characteristics. Also, the requested fee of €525.07 per month per pregnant woman is a strength of this study because of the applicability and thoughtfulness of this item. It is very likely that this price will actually cover the costs of a remote monitoring prenatal follow-up program. Finally, by adding this supplement to the RIZIV costs, there will be no increase in costs between the remote monitoring group and conventional care group in the three domains, but the prenatal follow-up and gestational outcomes will be improved for the remote monitoring group as we reported previously [10].

The main limitation of this study is its retrospective structure and the fact that the patients from the PREMOM study were not randomized. Nevertheless, the populations in the two arms were almost homogeneous regarding the baseline clinical characteristics. Second, the PREMOM study and this financial

analysis provides a picture of real-life practice in Belgium; we did receive the data from the patient files and the hospital bills, but we do not have information of the patients' hospital and medical consumption, or the patients' social costs (eg, transportation and travel costs and the cost of lost employment income for the time spent for in hospital visits). Our results could also differ in different HCSs and different economic and social settings, such as in other countries. Additionally, this study is limited to 6 weeks after delivery. It is generally known that neonates that need intensive care at the time of their delivery will have a higher impact on health care costs then neonates who do not need this care. These costs are mostly due to rehospitalizations, acute care visits, or further intensive care for the rest of the infant's life [15-19]. Further, we did not investigate the quality-adjusted life years (QALYs), which can be used as a generic measure of effectiveness. QALYs are a generic measure of disease burden, including both the quality and the quantity of the life lived, and it assesses the value of costs of medical interventions. To conclude, we evaluated only one type of remote monitoring follow-up program, which does not allow our results to be transferred to other proprietary technologies with varying transmission frequencies and methods of alert notifications.

Comparisons With Previous Trials

Only two studies are known to have performed a cost analysis of a remote monitoring follow-up program in women with high-risk pregnancies. Morrison et al [9] performed a cost-effectiveness evaluation of remote monitoring in patients diagnosed with preterm labor. An average reduced cost of US $14,459 per pregnancy using remote monitoring services was obtained when compared to usual care. This cost reduction was due to reduced costs in antepartum hospitalization and intensive care nursery [9]. The conclusions of this article are in line with our main findings. Also, the study of Buysse et al [11] matches our principal findings. They obtained a cost reduction of €145,882 per year for high-risk pregnancies. But, unlike our study, these researchers did not use real-life data from patients in a remote monitoring program: they made a simulation exercise for all high-risk pregnancies that would qualify for home monitoring.

Possible Explanations

The main objective of our study was to compare direct costs of a prenatal follow-up program for women diagnosed with GHD between remote monitoring and conventional care in hospital visits for a single-center population based on the initial assumption that remote monitoring technologies were provided with no additional costs. Early detection of clinical and device-related critical events provided by remote monitoring may have a positive impact on complication rates such as the development of severe hypertension, the need of inductions, prenatal hospitalizations, and neonatal hospitalizations. In our previously mentioned study, we reported a reduction in the prevalence of preeclampsia, hospitalization of the mother and the neonate, and inductions of labor [10]. In summary, by adding remote monitoring to the prenatal care of women at risk of these disorders, the risk of development of a severe hypertensive disorder is reduced and there are large potential benefits in terms of social and hospital expenditure restraint. These results can be read in Multimedia Appendix 5. In line with these benefits that are obtained with remote monitoring, the costs necessary for the medical care of the previously mentioned complications are reduced and/or avoided in the remote monitoring group and not in the conventional care group. The slightly higher costs of the medications for the patients of the remote monitoring group, when compared to conventional care group, can be explained by the higher need of medication for those patients. During the remote monitoring process, it is easy to make some changes in the antihypertensive treatment because their daily parameters are constantly at hand [10]. Women in the conventional care group will have less medication changes due to the lack of daily follow-up of their blood pressure.

The suggested €525.07 per month per pregnant woman fee for funding remote monitoring allows for HCS to not be elevated. By showing that there is no significant difference in costs between the remote monitoring group and conventional care group, a door is opened for policy makers charged with deciding how limited health care resources should be allocated in the era of exploding needs. This study, together with our previous report, states that better prenatal follow-up and gestational outcomes for the same cost as conventional care are possible by adding remote monitoring to the care of pregnant women with GHD.

Recommendations for Further Research

Firstly, it would also be useful to investigate the QALYs for both the mother and the neonate who received remote monitoring to make further recommendations about this topic. This study is also shortened to postnatal follow-up until 6 weeks after delivery. It would be interesting to monitor the neonates in both groups—remote monitoring and conventional care groups—for longer than 6 weeks postpartum to get insights into the long-term cost benefits. Lastly, because the social costs (eg, transportation and travel costs and the cost of lost employment income for the time spent for in hospital visits) are not taken into account, it would be interesting to make additional analyses with these type of costs included. It is plausible that the differences in costs will be even greater when the previously mentioned items are taken into account.

Conclusions

The results of this study show that a remote monitoring prenatal follow-up of women with GHD will not increase the costs for the HCS, RIZIV, or patient in comparison with conventional care. Furthermore, a RIZIV fee of €525.07 per month per pregnant woman allows the implementation of remote monitoring without increasing the health care costs for the remote monitoring group. These results are useful for policy makers charged with deciding how limited health care resources should be allocated in the era of exploding need. Further research of the long-term cost-effectiveness of remote monitoring, the QALYs, and social costs is recommended.

Acknowledgments

We would like to thank Mattias Neyt and the billing service of the Ziekenhuis Oost-Limburg (Genk, Belgium) for their assistance and support during this study. This study is part of the Limburg Clinical Research Program (LCRP) UHasselt-ZOL-Jessa, supported by the foundation Limburg Sterk Merk, province of Limburg, Flemish government, Hasselt University, Ziekenhuis Oost-Limburg and Jessa Hospital. This work was supported by Foundation Mustela (Laureate 2016).

Conflicts of Interest

None declared.

Multimedia Appendix 1

A detailed overview of the included costs.

[PDF File (Adobe PDF File), 39KB - jmir_v20i3e102_app1.pdf]

Multimedia Appendix 2

The influence of the maternal demographics and characteristics on the health care costs.

[PDF File (Adobe PDF File), 54KB - jmir_v20i3e102_app2.pdf]

Multimedia Appendix 3

Health care costs.

[PDF File (Adobe PDF File), 30KB - jmir_v20i3e102_app3.pdf]

Multimedia Appendix 4

Detailed information the mean time of prenatal RM follow-up/pregnant woman.

[PDF File (Adobe PDF File), 33KB - jmir_v20i3e102_app4.pdf]

Multimedia Appendix 5

Remote monitoring of hypertension diseases in pregnancy: a pilot study.

[PDF File (Adobe PDF File), 17MB - jmir_v20i3e102_app5.pdf]

References

1. Corwin MJ, Mou SM, Sunderji SG, Gall S, How H, Patel V, et al. Multicenter randomized clinical trial of home uterine activity monitoring: pregnancy outcomes for all women randomized. Am J Obstet Gynecol 1996 Nov;175(5):1281-1285. [Medline: 8942501]
2. Wapner RJ, Cotton DB, Artal R, Librizzi RJ, Ross MG. A randomized multicenter trial assessing a home uterine activity monitoring device used in the absence of daily nursing contact. Am J Obstet Gynecol 1995 Mar;172(3):1026-1034. [Medline: 7892843]
3. Homko CJ, Santamore WP, Whiteman V, Bower M, Berger P, Geifman-Holtzman O, et al. Use of an internet-based telemedicine system to manage underserved women with gestational diabetes mellitus. Diabetes Technol Ther 2007 Jun;9(3):297-306. [doi: 10.1089/dia.2006.0034] [Medline: 17561800]
4. Pérez-Ferre N, Galindo M, Fernández MD, Velasco V, Runkle I, de la Cruz MJ, et al. The outcomes of gestational diabetes mellitus after a telecare approach are not inferior to traditional outpatient clinic visits. Int J Endocrinol 2010;2010:386941 [FREE Full text] [doi: 10.1155/2010/386941] [Medline: 20628517]
5. Pérez-Ferre N, Galindo M, Fernández MD, Velasco V, de la Cruz MJ, Martín P, et al. A telemedicine system based on Internet and short message service as a new approach in the follow-up of patients with gestational diabetes. Diabetes Res Clin Pract 2010 Feb;87(2):e15-e17. [doi: 10.1016/j.diabres.2009.12.002] [Medline: 20044162]
6. Dalfrà MG, Nicolucci A, Lapolla A, TISG. The effect of telemedicine on outcome and quality of life in pregnant women with diabetes. J Telemed Telecare 2009;15(5):238-242. [doi: 10.1258/jtt.2009.081213] [Medline: 19590029]
7. Rauf Z, O'Brien E, Stampalija T, Ilioniu FP, Lavender T, Alfirevic Z. Home labour induction with retrievable prostaglandin pessary and continuous telemetric trans-abdominal fetal ECG monitoring. PLoS One 2011;6(11):e28129 [FREE Full text] [doi: 10.1371/journal.pone.0028129] [Medline: 22140522]
8. O'Brien E, Rauf Z, Alfirevic Z, Lavender T. Women's experiences of outpatient induction of labour with remote continuous monitoring. Midwifery 2013 Apr;29(4):325-331. [doi: 10.1016/j.midw.2012.01.014] [Medline: 23159160]

9. Morrison J, Bergauer NK, Jacques D, Coleman SK, Stanziano GJ. Telemedicine: cost-effective management of high-risk pregnancy. Manag Care 2001 Nov;10(11):42-46, 48. [Medline: 11761593]
10. Lanssens D, Vandenberk T, Smeets C, De Cannière H, Molenberghs G, Van Moerbeke A, et al. Remote monitoring of hypertension diseases in pregnancy: a pilot study. JMIR Mhealth Uhealth 2017;5(3):1. [Medline: 28279948]
11. Buysse H, De Moor G, Van Maele G, Baert E, Thienpont G, Temmerman M. Cost-effectiveness of telemonitoring for high-risk pregnant women. Int J Med Inform 2008 Jul;77(7):470-476. [Medline: 17923433]
12. Devlieger R, Martens E, Martens G, Van Mol C, Cammu H. Perinatale activiteiten in Vlaanderen 2015. Brussels: SPE; 2016.
13. Rijksinstituut voor ziekte- en invaliditeitsverzekering Brussel. 2017. URL: http://www.riziv.fgov.be/nl/Paginas/default.aspx [accessed 2018-02-28] [WebCite Cache ID 6xZLAb8zz]
14. Malehi A, Pourmotahari F, Angali KA. Statistical models for the analysis of skewed healthcare cost data: a simulation study. Health Econ Rev 2015;5(11):a. [Medline: 26029491]
15. Korvenranta E, Lehtonen L, Rautava L, Häkkinen U, Andersson S, Gissler M, PERFECT Preterm Infant Study Group. Impact of very preterm birth on health care costs at five years of age. Pediatrics 2010 May 05;125(5):e1109-e1114. [doi: 10.1542/peds.2009-2882] [Medline: 20368320]
16. Petrou S. Economic costs associated with moderate and late preterm birth: primary and secondary evidence. Semin Fetal Neonatal Med 2012;17(3):9. [Medline: 22364679]
17. Petrou S, Eddama O, Mangham L. A structured review of the recent literature on the economic consequences of preterm birth. Arch Dis Child Fetal Neonatal Ed 2011 May;96(3):F225-F232. [doi: 10.1136/adc.2009.161117] [Medline: 20488863]
18. Cavallo M, Gugiatti A, Fattore G, Gerzeli S, Barbieri D. Cost of care and social consequences of very low birth weight infants without premature-related morbidities in Italy. Ital J Pediatr 2015;41(59):1-17. [Medline: 26286526]
19. Cuevas K, Silver D, Brooten D, Youngblut J, Bobo CM. The cost of prematurity: hospital charges at birth and frequency of rehospitalizations and acute care visits over the first year of life: a comparison by gestational age and birth weight. Am J Nurs 2005 Jul;105(7):56-64; quiz 65 [FREE Full text] [Medline: 15995395]

Abbreviations

RM: remote monitoring
CC: conventional care
GHD: gestational hypertensive diseases
HCS: health care system
IQR: interquartile range
PREMOM study: Pregnancy Remote Monitoring study
QALY: quality-adjusted life years
RIZIV: National Institution for Insurance of Disease and Disability [Rijksinstituut voor Ziekte- en Invaliditeitverzekering]

Edited by G Eysenbach; submitted 02.12.17; peer-reviewed by I Kedan, K Deldar, S Ben Omrane; comments to author 10.01.18; revised version received 23.01.18; accepted 23.01.18; published 26.03.18.

Please cite as:
Lanssens D, Vandenberk T, Smeets CJP, De Cannière H, Vonck S, Claessens J, Heyrman Y, Vandijck D, Storms V, Thijs IM, Grieten L, Gyselaers W
Prenatal Remote Monitoring of Women With Gestational Hypertensive Diseases: Cost Analysis
J Med Internet Res 2018;20(3):e102
URL: http://www.jmir.org/2018/3/e102/
doi:10.2196/jmir.9552
PMID:29581094

Original Paper

Mode Equivalence of Health Indicators Between Data Collection Modes and Mixed-Mode Survey Designs in Population-Based Health Interview Surveys for Children and Adolescents: Methodological Study

Elvira Mauz[1], MPH, Dipl-Psych; Robert Hoffmann[1], MA; Robin Houben[1], MSc EPH; Laura Krause[1], MA, Dipl Pad, Dr Phil; Panagiotis Kamtsiuris[1], Dipl Econ; Antje Gößwald[1], MSc Epi, Dr Med

Department of Epidemiology and Health Monitoring, Robert Koch Institute, Berlin, Germany

Corresponding Author:
Elvira Mauz, MPH, Dipl-Psych
Department of Epidemiology and Health Monitoring
Robert Koch Institute
PO Box 650261
Berlin, 13302
Germany
Phone: 49 30 18754 3332
Fax: 49 30 18754 3449
Email: MauzE@rki.de

Abstract

Background: The implementation of an Internet option in an existing public health interview survey using a mixed-mode design is attractive because of lower costs and faster data availability. Additionally, mixed-mode surveys can increase response rates and improve sample composition. However, mixed-mode designs can increase the risk of measurement error (mode effects).

Objective: This study aimed to determine whether the prevalence rates or mean values of self- and parent-reported health indicators for children and adolescents aged 0-17 years differ between self-administered paper-based questionnaires (SAQ-paper) and self-administered Web-based questionnaires (SAQ-Web), as well as between a single-mode control group and different mixed-mode groups.

Methods: Data were collected for a methodological pilot of the third wave of the "German Health Interview and Examination Survey for Children and Adolescents". Questionnaires were completed by parents or adolescents. A population-based sample of 11,140 children and adolescents aged 0-17 years was randomly allocated to 4 survey designs—a single-mode control group with paper-and-pencil questionnaires only (n=970 parents, n=343 adolescents)—and 3 mixed-mode designs, all of which offered Web-based questionnaire options. In the concurrent mixed-mode design, both questionnaires were offered at the same time (n=946 parents, n=290 adolescents); in the sequential mixed-mode design, the SAQ-Web was sent first, followed by the paper questionnaire along with a reminder (n=854 parents, n=269 adolescents); and in the preselect mixed-mode design, both options were offered and the respondents were asked to request the desired type of questionnaire (n=698 parents, n=292 adolescents). In total, 3468 questionnaires of parents of children aged 0-17 years (SAQ-Web: n=708; SAQ-paper: n=2760) and 1194 questionnaires of adolescents aged 11-17 years (SAQ-Web: n=299; SAQ-paper: n=895) were analyzed. Sociodemographic characteristics and a broad range of health indicators for children and adolescents were compared by survey design and data collection mode by calculating predictive margins from regression models.

Results: There were no statistically significant differences in sociodemographic characteristics or health indicators between the single-mode control group and any of the mixed-mode survey designs. Differences in sociodemographic characteristics between SAQ-Web and SAQ-paper were found. Web respondents were more likely to be male, have higher levels of education, and higher household income compared with paper respondents. After adjusting for sociodemographic characteristics, only one of the 38 analyzed health indicators showed different prevalence rates between the data collection modes, with a higher prevalence rate for lifetime alcohol consumption among the online-responding adolescents (P<.001).

Conclusions: These results suggest that mode bias is limited in health interview surveys for children and adolescents using a mixed-mode design with Web-based and paper questionnaires.

(J Med Internet Res 2018;20(3):e64) doi:10.2196/jmir.7802

KEYWORDS

public health; child and adolescent health; health surveys; survey methods; mixed-mode survey; paper-and-pencil questionnaire; online questionnaire; mode effects

Introduction

The assessment of population health using health interview surveys is an established method in many countries and is a cornerstone of health reporting, health policies, and health sciences. However, epidemiological studies have shown decreasing response rates since the 1990s [1-3]. The use of mixed-mode health interview surveys offers respondents various data collection modes and can increase the response rate, improve sample composition, and reduce overall costs [3,4]. Currently, there is considerable interest in using Web-based health survey interviews because of lower costs and faster data availability. Web-based surveys are increasingly becoming standard [5], and they are frequently combined with other modes in mixed-mode designs [6]. However, the use of different survey modes may increase the risk of measurement error (mode effects) [5].

Mode effects are systematic distortions caused by different survey modes or interview situations [5]. They often arise when there are large methodological differences in the survey situation (self-administered questionnaire vs interviews) or the communication channel (auditory vs visual) [3]. Such differences are minimal between self-administered paper-based questionnaires (SAQ-paper) and self-administered Web-based questionnaires (SAQ-Web)—both are conducted without an interviewer and both use visual perception. For this reason, these 2 self-administered modes (SAQ-Web and SAQ-paper) are considered mode equivalent [4,7,8]. Mode equivalence is shown if an individual gives the same response to the same question or instrument administered through 2 different modes, leading to the same results [9]. For example, research has shown no differences between the 2 data collection modes in prevalence rates of diseases among adult populations [10,11] or in reported health behaviors among adolescents [12].

However, researchers have discussed mode effects for sensitive topics. Web-based responses are associated with both anonymity and greater individualization. Consequently, SAQ-Web participants are not affected by social desirability; rather, they are less orientated toward social norms. Therefore, SAQ-Web mode yields the most honest reports, especially compared with interview modes [13,14]. Furthermore, differences have been found between the 2 self-administered modes, for example, in political attitudes [15], reporting of sensitive sexual behaviors [16], or adolescent risk behavior [17]. However, there is high consistency of responses across modes, with only a few respondents taking advantage of the greater privacy of the Web mode [16]. Hence, possible mode effects should be investigated before changing or adding modes to existing health surveys. In ongoing longitudinal studies, changing the mode or offering a second mode may risk time-based comparability.

The German Health Interview and Examination Survey for Children and Adolescents (KiGGS) is a nationally representative health interview and examination survey of children and adolescents in Germany [18,19]. It is part of the nationwide health monitoring system administered by the German national public health institute (Robert Koch Institute) [20,21]. KiGGS obtains representative cross-sectional information on German children and adolescents aged 0-17 years at regular intervals. Additionally, based on the first cross-sectional sample (KiGGS baseline; 2003-2006), a KiGGS cohort has been implemented. The baseline respondents are being followed throughout their life course into adulthood [21]. The survey involves physical examinations and tests, as well as laboratory analysis of urine and blood parameters. All the parents and adolescents aged 1117 years completed paper-based questionnaires [20]. The first follow-up, KiGGS Wave 1 (2009-2012), was conducted using telephone interviews of parents and adolescents [22]. KiGGS Wave 2 (2014-2017) involved a health interview and examination, continuing the baseline concept [23]. The aim of the KiGGS survey is to provide current data on population health, health determinants, and the utilization of health care services. In addition, information is gathered about the incidence of disorders as well as trajectories of multiple health indicators throughout the life course. The data are widely used in national health reporting, health policies, and public health research.

When planning population-based (health) studies like KiGGS, the survey design must minimize total survey error [24,25]. In addition to lower data quality owing to measurement errors such as mode effects, the total survey error comprises different kinds of systematic errors—an undervalued sample size leads to imprecise estimates (sampling error) and the composition of the sample might be different from the target population (coverage error) owing to errors in the sampling procedure or because of systematic nonresponse (nonresponse bias). All these aspects were examined in a methodological pilot study as part of the KiGGS Wave 2 pretest. The pilot study aimed to compare 3 mixed-mode survey designs using Web- and paper-based questionnaires with a single-mode SAQ-paper design in terms of response rates, sample composition, data quality, and effort [26]. The study also explored whether estimates of health indicators differed among the survey designs and data collection modes. This study focused only on the second aim of the pilot study and addressed 2 research questions:

- Are there any differences in the prevalence rates or mean values of core public health indicators for children and adolescents aged 0-17 years between the single-mode control group using only SAQ-paper and different mixed-mode groups that combine offers of SAQ-paper and SAQ-Web?
- Are there any differences in prevalence rates or mean values of these indicators between the 2 data collection modes (SAQ-paper and SAQ-Web) if all online respondents are

pooled and all paper-and-pencil respondents are pooled across all survey designs?

Methods

Study Design

The methodological pilot study used a sample of children and adolescents registered in the local resident registries of 20 municipalities in 5 federal states of Germany, covering urban and rural areas as well as the eastern and western regions of the country.

Data were collected using SAQ-Web or SAQ-paper methods. All selected individuals were invited by mail to participate in the study. They were sent a cover letter with the invitation to participate, information about the study and data privacy, and an informed consent form. Depending on the allocated mode, the invitation comprised a username and password for participation through the Web option along with a paper questionnaire for those allocated to the concurrent mixed-mode design, only a paper questionnaire in the single-mode design, or only the access data for the online questionnaire in the sequential mixed-mode design. The SAQ-Web questionnaire was only optimized for desktop computers. A reminder was sent by mail to respondents who had not replied within 3 weeks of the initial invitation. Participants who did not respond to the reminder were telephoned up to 5 times 4 weeks after the initial invitation. As an additional motivation for prospective participants, each parent and adolescent who had completed a questionnaire received a shopping voucher to the value of €10. The methodological pilot study strictly adhered to the data protection regulations set out in the German Federal Data Protection Act. Participation in the study was voluntary. All parents and participating adolescents were informed about the study's aims and content, as well as data protection, and they provided informed consent. Following the strict data privacy protocol, prospective participants between the ages of 11 and 17 years received their questionnaires only after their parents provided consent.

Different questionnaires were used for different age groups. Main health indicators were included on the health questionnaires for parents of all age groups (0-17 years), and self-report data for main health indicators were obtained from adolescents aged 11-17 years. To reduce the risk of mode effects, the 2 questionnaires were designed to be as similar as possible and contained the same wording for the questions and response categories. On the basis of the unified-mode design [27], the wording and formatting of questions and response categories were standardized. To help participants visually distinguish single-choice questions from multiple-choice questions, all survey modes used the same checkbox design. Single-choice checkboxes were round, whereas multiple-choice checkboxes were rectangular. Additionally, multiple-choice questions included the instruction "Multiple entries are possible." For filter questions, Web-based questionnaires were optimized with filter skips whenever the perceivability of the questions was not impaired. Plausibility checks and ranges were defined for the Web-based questionnaire. Additionally, soft prompting was programmed into the Web-based questionnaire to reduce item nonresponse. These differences were used to capitalize on the advantages of the Web mode for better data quality, and they were the only mode-specific design differences. Detailed information of the survey design and other technical aspects of the Web-based part of the survey are described in a "Checklist for Reporting Results of Internet E-Surveys" [28] (Multimedia Appendix 1).

As shown in Figure 1, a gross sample of 11,140 children and adolescents was randomly allocated to four survey designs:

1. A single-mode survey design as a control group—respondents were sent an invitation letter and paper-and-pencil questionnaires, followed by a reminder after 3 weeks
2. A sequential mixed-mode survey design—respondents were sent an invitation letter and an online access code, followed 3 weeks later with a reminder letter and a paper-based questionnaire
3. A concurrent mixed-mode survey design—respondents were sent an invitation letter, a paper-based questionnaire, and an online access code (a longer version of the questionnaire was tested with a subgroup of the concurrent mixed-mode design, but this subgroup was excluded from this study) and
4. A preselect mixed-mode design—respondents were sent the invitation along with a postcard asking participants to choose one of the 2 options (SAQ-Web or SAQ-paper), followed by a reminder with the same offer

There were no statically significant differences in the (gross) sample composition across the 4 design groups in terms of known sample characteristics, such as age, sex, municipality size, region, or respondent citizenship, which were obtained from local registries.

The combined response rate for all survey designs was 38.43% (n=4032), following the internationally used Standard Definitions of Outcome Rates for Surveys of the American Association for Public Opinion Research (AAPOR Response Rate 2) [29]. There were no significant differences in response rates among the concurrent mixed-mode design, the sequential mixed-mode design, and the single-mode control group design. However, there was a significantly lower response rate in the preselect mixed-mode design. Detailed comparisons of response rates, sample compositions, data quality, and efforts among the different survey designs have been published previously [26].

Database

For this study, only survey design groups using the same version of the questionnaire were included, with 3468 completed parent-reported health questionnaires for children and adolescents aged 0-17 years and 1194 questionnaires completed by adolescents aged 11-17 years. A response was defined as one completed health questionnaire from either parents or children. Hence, a valid response did not require both parents and children to complete all requested questionnaires. To answer the first research question regarding mode equivalence across the different survey designs, we compared the single-mode control group with each of the 3 mixed-mode groups. To answer the second research question regarding mode equivalence

between the 2 data collection modes, data from all survey designs were pooled (Table 1).

Sociodemographic Characteristics of Responding Parents and Adolescents by Survey Design and Data Collection Mode

Analyzed Sociodemographic Characteristics

The sample compositions of participating parents and adolescents were described by various sociodemographic characteristics separately by survey design and data collection mode. The variables examined included individual adolescent characteristics (age, sex, migration background, and highest level of education reached or aspired); parental characteristics (age, marital status, and participating parent); location (municipality size and region [East vs West Germany]); and household properties (education level and net household income). Household education level was measured using the Comparative Analysis of Social Mobility in Industrial Nations [30]. Household income was assessed using a question on household monthly net income.

Figure 1. Study design of the methodological pilot study. SAQ-paper: self-administered paper-based questionnaire; SAQ-Web: self-administered Web-based questionnaire.

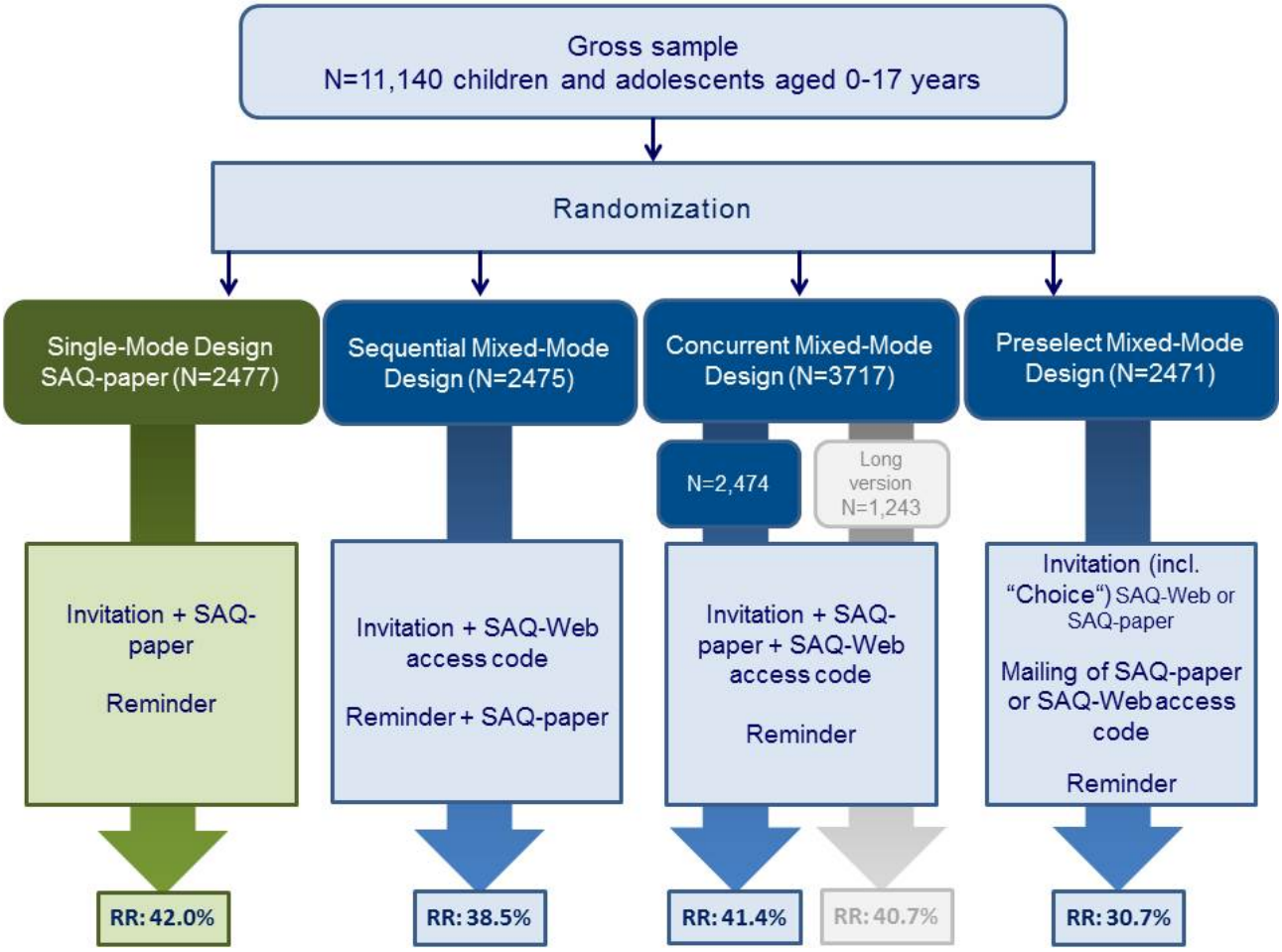


Table 1. Cases used in this study.

Questionnaire type	Survey design										Mode group		
	Single-mode design	Sequential mixed-mode design			Concurrent mixed-mode design			Preselect mixed-mode design					
	SAQ-paper[a]	SAQ-paper	SAQ-Web[b]	Total	SAQ-paper	SAQ-Web	Total	SAQ-paper	SAQ-Web	Total	SAQ-paper	SAQ-Web	Total
Parent-completed health questionnaires	970	488	366	854	837	109	946	465	233	698	2760	708	3468
Adolescent-completed health questionnaires	343	117	152	269	244	46	290	191	101	292	895	299	1194

[a]SAQ-paper: self-administered paper-based questionnaire.

[b]SAQ-Web: self-administered Web-based questionnaire.

Statistical Methods

Differences between the control group and the different mixed-mode groups and between the 2 data collection modes were tested using chi-squared tests.

Mode Equivalence of Health Indicators Between Survey Designs and Data Collection Modes

A wide range of health status indicators and health behaviors for children and adolescents with high public health relevance were analyzed to identify differences between the mixed-mode designs and the single-mode control group, as well as mode differences between SAQ-paper and SAQ-Web.

Analyzed Indicators of Physical and Mental Health

Lifetime diagnoses of asthma, hay fever, atopic eczema, and attention-deficit hyperactivity disorder (ADHD) were indicated by parents. Recurrent pain during the last 3 months was measured using the adolescents' self-reports. Self-rated health (SRH) and chronic diseases were evaluated by parental report using the Minimum European Health Module questions [31], modified for children. Adolescents also answered the SRH question. Impairments owing to health problems were evaluated with a question from the Children with Special Health Care Needs Screener, which was answered by parents [32]. To define obesity, body mass index was calculated based on self-reported weight and height for adolescents and parent-reported weight and height for children aged 3-10 years. The body mass index cut-offs used in this study were determined by German norms [33].

Child and adolescent mental health problems were evaluated using the parent- and self-report Strengths and Difficulties Questionnaire (SDQ) [34]. An SDQ total difficulties score was calculated for all children and adolescents. Participants with a borderline or abnormal score (based on German norms) [35] were defined as at risk for emotional and behavioral symptoms. Participants with borderline or abnormal SDQ impact scores were defined as at risk for psychosocial impairment.

Analyzed Indicators of Health Care Utilization

As indicators of health care use, pediatrician and orthodontist visits during the past 12 months for adolescents and parent-reported visits to any doctor for children under 11 years were analyzed [36].

Analyzed Measure of Health-Related Quality of Life

Health-related quality of life (HRQoL) was measured using KIDSCREEN-27 for adolescents aged 11-17 years, with 5 subscores for physical and psychological well-being, relationships with peers and parents, and school well-being. Scores were summed and transformed into *t* values [37].

Analyzed Health Behaviors

Adolescents reported their current smoking status, water pipe consumption during the past 12 months, second-hand smoke exposure [38], and lifetime consumption and current use of screen-based media. Excessive use of screen-based media was defined as more than 2 hours per day [39]. Harmful alcohol use and binge drinking were defined using responses to the Alcohol Use Disorders Identification Test (AUDIT-C) [40].

Following the recommendation of the World Health Organization [41], healthy physical activity was defined as physical activity for at least 60 min per day. Low physical activity was defined as less than 2 days per week of at least 60 min of activity. All questions on physical activity were answered by adolescents aged 11-17 years.

Statistical Methods

We calculated prevalence rates for dichotomous health indicators and mean values for HRQoL (a scale outcome) by survey design and data collection mode. We compared these values using *z* or *t* tests.

Due to the different sample compositions of the SAQ-paper and SAQ-Web groups (see the Results), it was necessary to control for sociodemographic characteristics to identify possible mode effects. Survey modes can differ in selection (different population groups prefer different modes) and measurement (different answers are given by the same person under different modes of administration), so these differences are confounded [42]. Additionally, health status and health behavior differ by sex, education, and other sociodemographic characteristics [43,44]. To eliminate the risk of confounding, we adjusted for sociodemographic characteristics by calculating adjusted prevalence rates using predictive margins [45] based on logistic or linear regression models with sociodemographic factors as covariates. To analyze indicators based on parental reports, we included child attributes (age, sex, and migration background); parental attributes (relationship to the child, age, and marital status); household attributes (education and income); and regional attributes (region and municipality size). Adolescents' reports were adjusted by child attributes, including the highest level of education completed, as well as household attributes and location. The mode of data collection was another covariate used to identify adjusted prevalence for each mode. Differences were tested using *z* or *t* tests.

For the survey design comparison, crude as well as adjusted prevalence rates and mean values were calculated. A statistical test for diversity was conducted between the single-mode control group design and each of the 3 mixed-mode designs. Because the survey design samples did not differ in sociodemographic characteristics (see the Results) and there were only marginal differences between the 2 approaches, only the results for crude prevalence rates or mean values without adjustment for sociodemographic characteristics to simplify the presentation of results are shown here.

Handling of Multiple Testing

In total, we analyzed 12 health indicators using the parental sample and 28 using the adolescent sample. For these health indicators, we tested each mixed-mode survey design against the control group. Additionally, we used 2 other statistical tests to identify differences between the data collection modes, using first the crude values and then the adjusted values.

Regarding the research questions, a sensitive approach to detect possible differences (ie, a higher probability of accepting the null hypothesis) is needed. Therefore, we decided to address the statistical problem of multiple testing by correcting the significance level only for the number of tests performed for

each health indicator. This was done only for tests comparing the different survey designs. We used the Bonferroni correction method to neutralize the accumulation of α-error [46], using an adjusted significance level of $P<.02$ to examine differences between the mixed-mode survey designs and the single-mode control group. For the comparison of data collection modes, a significance level of $\alpha=.05$ was used.

Results

Sociodemographic Characteristics of Responding Parents and Adolescents in Different Survey Designs and Different Data Collection Modes

Responding Parents

There were no statistically significant differences in sample composition between the mixed-mode survey designs and the single-mode control group for participating parents. However, the sample sociodemographic characteristics differed significantly between data collection modes (Multimedia Appendix 2). Parents responding online were more often married and had higher household education levels, higher incomes, and younger children than those who responded to the SAQ-paper. More fathers responded via the Web-based questionnaire than in the paper-and-pencil group. There were no significant differences in migration background, parental age. *P* values close to significant level are found for region of residence ($P=.08$), municipally size ($P=.05$) or child's sex ($P=.06$).

Responding Children and Adolescents

For the responding children and adolescents (aged 11-17 years), there were no statistically significant differences in sociodemographic characteristics between the different survey designs, but adolescents responding online were more often male, had reached or aspired to reach higher levels of education, and were more likely to live in households with higher education and higher income, compared with adolescents who responded to the SAQ-paper (Multimedia Appendix 3).

Mode Equivalence of Health Indicators Between Survey Designs and Data Collection Modes

Physical and Mental Health

The analyzed indicators of physical and mental health status showed no statistically significant differences by survey design or data collection mode (Table 2). Across modes and designs, parents reported the same results for SRH, chronic disease, impairment owing to health problems, lifetime prevalence of diagnosed diseases, obesity, and mental health problems and impairment. Adolescent self-reports showed no statistically significant differences in SRH, mental health problems and impairment, or chronic pain.

Health Care Utilization

No differences in the crude or adjusted prevalence rates were found in adolescent-reported 12-month use of pediatric or orthodontic services (Table 3). The crude prevalence of parent-reported 12-month use of any doctor and of pediatric services (for children under 11 years) differed significantly, with more frequent reports of doctor's visits in the SAQ-Web group. After adjusting for sociodemographic attributes, this difference disappeared. There were no significant differences between the mixed-mode design groups and the control group for any of the analyzed indicators of health care utilization.

Health-Related Quality of Life

HRQoL, measured using the 5 dimensions of the KIDSCREEN-27 for adolescents, was the only indicator scale analyzed. Independent of adjustment, there were no significant differences between the 2 data collection modes (SAQ-paper and SAQ-Web) for any of the observed dimensions (Table 4). Regarding survey design, better psychological well-being was reported in the concurrent mixed-mode design and better relations with parents were reported in the preselect mixed-mode survey design, compared with the single-mode control group. After correcting the significance level for multiple testing, no differences were found by survey design.

Health Behaviors

The crude prevalence of lifetime alcohol consumption (self-reported by adolescents aged 11-17 years), as well as hazardous consumption and binge drinking (based on AUDIT-C reports), showed significant differences between SAQ-paper and SAQ-Web, with higher levels of alcohol consumption reported by online participants (Table 5). Although the differences in hazardous consumption and binge drinking between the 2 modes of data collection disappeared after controlling for sociodemographic characteristics, significantly more online respondents than paper-and-pencil respondents reported that they had consumed alcohol.

There were no differences in other health behaviors assessed (tobacco consumption, physical activity, and media consumption) by survey design or data collection mode.

Table 2. Physical and mental health status of children and adolescents aged 0-17 years by survey design and data collection mode (prevalence rates).

Physical and mental health status	Survey design[a]							Data collection mode[a]					
	Single-mode design	Sequential MM[b] design		Concurrent MM[b] design		Preselect MM[b] design		SAQ-Web[c] (crude)	SAQ-paper[d] (crude)		SAQ-Web[c] (adjusted[e])	SAQ-paper[d] (adjusted[e])	
	n (%)	n (%)	*P* [f]	n (%)	*P* [f]	n (%)	*P* [f]	n (%)	n (%)	*P*	n (%)	n (%)	*P*
General health status													
Self-rated health (very good, good)[g]	964 (97.6)	848 (97.1)	.46	942 (97.3)	.71	693 (97.3)	.65	708 (97.0)	2739 (97.41)	.60	708 (96.7)	2739 (97.48)	.31
Self-rated health (very good, good)[h]	338 (90.2)	267 (92.1)	.41	286 (87.1)	.22	290 (92.8)	.26	299 (90.6)	881 (90.5)	.93	299 (90.1)	881 (90.6)	.81
Chronic disease (yes)[g]	962 (10.2)	849 (10.7)	.71	941 (7.4)	.03[e]	691 (9.6)	.67	708 (10.2)	2735 (9.25)	.47	708 (10.5)	2735 (9.18)	.31
Impairment owing to health problems (yes)[g]	956 (3.2)	851 (4.0)	.39	937 (3.4)	.83	692 (3.6)	.69	708 (4.1)	2728 (3.41)	.40	708 (4.5)	2728 (3.36)	.19
Allergies													
Bronchial asthma (lifetime diagnosis)[g]	951 (4.5)	838 (6.1)	.14	928 (4.6)	.91	675 (5.2)	.54	701 (4.6)	2691 (5.20)	.48	701 (5.0)	2691 (5.10)	.91
Neurodermatitis (lifetime diagnosis)[g]	951 (15.0)	839 (16.0)	.59	927 (13.8)	.45	678 (17.4)	.20	704 (17.0)	2691 (14.98)	.19	704 (17.5)	2691 (14.95)	.12
Hay fever (lifetime diagnosis)[g]	955 (11.2)	831 (11.2)	.99	931 (10.7)	.75	680 (11.5)	.87	702 (11.3)	2695 (11.09)	.91	702 (12.0)	2695 (10.96)	.50
Chronic pain													
Headache (recurrent during the last 3 months)[h]	331 (35.6)	260 (37.7)	.61	274 (33.9)	.66	282 (31.2)	.24	299 (32.1)	847 (35.4)	.30	299 (33.8)	847 (34.8)	.77
Dorsal pain (recurrent during the last 3 months)[h]	321 (25.9)	257 (23.0)	.42	265 (24.2)	.64	276 (27.5)	.64	298 (22.5)	820 (26.2)	.19	298 (23.6)	820 (25.8)	.45
Any pain (recurrent during the last 3 months)[h]	343 (84.3)	265 (83.0)	.68	288 (81.9)	.44	291 (81.1)	.30	294 (82.3)	892 (82.7)	.87	294 (82.6)	892 (82.6)	.99
Mental health problems													
Attention-deficit hyperactivity disorder (lifetime diagnosis)[g]	824 (5.1)	716 (4.1)	.33	794 (5.0)	.96	593 (3.2)	.07	595 (3.5)	2332 (4.67)	.19	595 (3.9)	2332 (4.56)	.45
At risk for emotional and behavioral symptoms[g]	788 (13.6)	695 (15.5)	.29	754 (13.7)	.96	577 (12.7)	.62	598 (13.0)	2216 (14.12)	.49	598 (13.9)	2216 (13.89)	.97
At risk for emotional and behavioral symptoms[h]	331 (9.7)	255 (12.2)	.34	278 (14.0)	.10	283 (14.1)	.09	293 (13.7)	853 (12.0)	.46	293 (13.)	853 (11.9)	.41
At risk for impairment following psychosocial problems[g]	820 (18.5)	719 (17.4)	.56	789 (17.2)	.50	592 (16.6)	.33	600 (16.3)	2320 (17.80)	.39	600 (17.5)	2320 (17.48)	.99
At risk for impairment following psychosocial problems[h]	339 (16.5)	263 (19.0)	.43	286 (16.4)	.98	288 (18.1)	.61	298 (19.8)	877 (16.5)	.21	298 (19.5)	877 (16.6)	.30
Obesity													

Physical and mental health status	Survey design[a]							Data collection mode[a]					
	Single-mode design	Sequential MM[b] design		Concurrent MM[b] design		Preselect MM[b] design		SAQ-Web[c] (crude)	SAQ-paper[d] (crude)		SAQ-Web[c] (adjusted[e])	SAQ-paper[d] (adjusted[e])	
	n (%)	n (%)	*P* [f]	n (%)	*P* [f]	n (%)	*P* [f]	n (%)	n (%)	*P*	n (%)	n (%)	*P*
Obesity of children (aged 0-10 years)[g]	401 (2.7)	407 (2.7)	.97	423 (2.4)	.73	302 (4.0)	.38	370 (1.9)	1163 (3.18)	.14	370 (2.7)	1163 (3.49)	.46
Obesity of adolescents (aged 11-17 years)[h]	330 (5.2)	260 (4.6)	.76	275 (4.4)	.65	275 (3.6)	.36	291 (3.4)	848 (4.8)	.28	291 (4.1)	848 (4.6)	.78

[a]Sample sizes are shown in Table 1.

[b]MM: mixed-mode.

[c]SAQ-Web: self-administered Web-based questionnaire.

[d]SAQ-paper: self-administered paper-based questionnaire.

[e]Adjusted for age of the child (adolescent, parent); sex of the child (adolescent); relationship to the child (parent); household income (adolescent, parent); parental education (adolescent, parent); adolescent education (adolescent); region (adolescent, parent); municipality size (adolescent, parent); and parental marital status (parent).

[f]Tested against single-mode control group.

[g]Proxy-reported by parents of children and adolescents aged 0-17 years.

[h]Self-reported by adolescents aged 11-17 years.

Table 3. Health care utilization among children and adolescents aged 0-17 years by survey design and data collection mode (prevalence rates).

Health care utilization	Survey design[a]							Data collection mode[a]					
	Single-mode design	Sequential MM[b] design		Concurrent MM[b] design		Preselect MM[b] design		SAQ-Web[c] (crude)	SAQ-paper[d] (crude)		SAQ-Web[c] (adjusted[e])	SAQ-paper[d] (adjusted[e])	
	n (%)	n (%)	*P* [f]	n (%)	*P* [f]	n (%)	*P* [f]	n (%)	n (%)	*P*	n (%)	n (%)	*P*
Medical care use													
Any doctor (children aged 0-13 years; past 12 months)[g]	970 (91.3)	804 (92.7)	.31	935 (91.8)	.74	645 (91.5)	.93	594 (95.3)	2760 (91.05)	<.001	594 (92.2)	2760 (91.69)	.75
Pediatric services (children aged 0-13 years; past 12 months)[g]	970 (70.0)	803 (71.6)	.46	935 (72.0)	.34	645 (71.6)	.48	593 (80.9)	2760 (69.16)	<.001	593 (70.9)	2760 (71.47)	.75
Pediatric services (adolescents aged 14-17 years; past 12 months)[h]	205 (35.6)	147 (35.4)	.96	155 (29.0)	.18	159 (35.8)	.96	164 (34.1)	502 (34.1)	.99	164 (33.8)	502 (34.2)	.92
Orthodontic services (adolescents aged 14-17 years; past 12 months)[h]	332 (40.4)	260 (37.3)	.45	281 (39.5)	.83	288 (44.4)	.31	295 (39.7)	865 (40.7)	.75	295 (40.3)	865 (40.5)	.96

[a]Sample sizes are shown in Table 1.

[b]MM: mixed-mode.

[c]SAQ-Web: self-administered Web-based questionnaire.

[d]SAQ-paper: self-administered paper-based questionnaire.

[e]Adjusted for age of the child (adolescent, parent); sex of the child (adolescent); relationship to the child (parent); household income (adolescent, parent); parental education (adolescent, parent); adolescent education (adolescent); region (adolescent, parent); municipality size (adolescent, parent); and parental marital status (parent).

[f]Tested against single-mode control group.

[g]Proxy-reported by parents of children and adolescents aged 0-17 years.

[h]Self-reported by adolescents aged 11-17 years.

Table 4. Health-related quality of life of adolescents aged 11-17 years by survey design and data collection mode (mean values).

Health related quality of life	Survey design[a]							Data collection mode[a]					
	Single-mode design	Sequential MM[b] design		Concurrent MM[b] design		Preselect MM[b] design		SAQ-Web[c] (crude)	SAQ-paper[d] (crude)		SAQ-Web[c] (adjusted[e])	SAQ-paper[d] (adjusted[e])	
	n (%)	n (%)	*P* [h]	n (%)	*P* [h]	n (%)	*P* [h]	n (%)	n (%)	*P*	n (%)	n (%)	*P*
Dimensions													
Physical well-being (mean)[f]	337 (49.9)	263 (49.9)	.98	282 (49.2)	.38	290 (49.0)	.21	297 (49.5)	874 (49.5)	.95	297 (49.0)	874 (49.7)	.26
Psychological well-being (mean)[f]	336 (51.2)	265 (50.0)	.14	283 (49.6)	.05[g]	291 (49.8)	.07	297 (49.6)	877 (50.4)	.20	297 (49.3)	877 (50.5)	.07
Relations with parents (mean)[f]	331 (53.5)	263 (53.0)	.52	283 (53.3)	.74	290 (52.0)	.04[g]	296 (52.5)	871 (53.1)	.31	296 (52.3)	871 (53.2)	.18
Relations with peers (mean)[f]	340 (51.2)	266 (50.1)	.14	288 (50.5)	.37	292 (50.1)	.13	297 (49.8)	888 (50.8)	.12	297 (49.9)	888 (50.7)	.18
Well-being in school (mean)[f]	335 (51.9)	265 (51.6)	.66	283 (50.8)	.08	285 (51.4)	.43	294 (51.1)	873 (51.60)	.31	294 (50.9)	873 (51.70)	.17

[a]Sample sizes are shown in Table 1.

[b]MM: mixed-mode.

[c]SAQ-paper: self-administered paper-based questionnaire.

[d]SAQ-Web: self-administered Web-based questionnaire.

[e]Adjusted for age of the child (adolescent, parent); sex of the child (adolescent); relationship to the child (parent); household income (adolescent, parent); parental education (adolescent, parent); adolescent education (adolescent); region (adolescent, parent); municipality size (adolescent, parent); and parental marital status (parent).

[f]Self-reported by adolescents aged 11-17 years.

[g]Not significant, *P* value adjusted with Bonferroni correction.

[h]Tested against single-mode control group.

Table 5. Health behaviors of adolescents aged 11-17 years by survey design and data collection mode (prevalence-rates).

Health related quality of life	Survey design[a]							Data collection mode[a]					
	Single-mode design	Sequential MM[b] design		Concurrent MM[b] design		Preselect MM[b] design		SAQ-Web[c] (crude)	SAQ-paper[d] (crude)		SAQ-Web[c] (adjusted[e])	SAQ-paper[d] (adjusted[e])	
	n (%)	n (%)	*P* [i]	n (%)	*P* [i]	n (%)	*P* [i]	n (%)	n (%)	*P*	n (%)	n (%)	*P*
Tobacco consumption													
Current smoking status (yes)[f]	342 (8.8)	264 (8.3)	.85	288 (11.5)	.27	291 (10.7)	.43	295 (11.5)	889 (9.2)	.27	295 (11.6)	889 (9.2)	.25
Water pipe consumption (past 12 months, yes)[f]	339 (18.0)	262 (17.9)	.99	286 (18.2)	.95	291 (16.2)	.54	295 (20.3)	882 (16.7)	.17	295 (19.1)	882 (17.1)	.38
Second-hand smoke exposure (yes)[f]	308 (13.6)	239 (15.5)	.55	249 (10.8)	.31	259 (14.3)	.82	261 (11.1)	793 (14.4)	.16	261 (12.9)	793 (13.8)	.74
Alcohol consumption													
Lifetime consumption of alcohol (yes)[f]	343 (51.0)	262 (56.1)	.21	290 (50.7)	.93	292 (54.8)	.34	295 (61.4)	891 (50.3)	.001	295 (60.1)	891 (50.7)	<.001
Hazardous alcohol consumption (based on AUDIT-C[g])[f]	328 (12.2)	255 (12.5)	.90	284 (10.6)	.53	284 (12.7)	.86	293 (17.1)	857 (10.3)	.01	293 (14.9)	857 (11.9)	.16
Binge drinking (based on AUDIT-C)[f]	339 (7.7)	262 (7.7)	.99	288 (5.6)	.29	291 (8.9)	.57	295 (10.5)	884 (6.5)	.04	295 (10.0)	884 (7.2)	.13
Physical activity													
Physical activity consistent with WHO[h] guidelines[f]	343 (6.1)	265 (5.3)	.66	288 (3.1)	.07	292 (3.8)	.17	295 (4.1)	892 (4.8)	.58	295 (3.3)	892 (5.2)	.16
Low physical activity[f]	343 (14.9)	265 (12.5)	.39	288 (15.3)	.89	292 (17.1)	.44	295 (13.9)	892 (15.2)	.57	295 (14.5)	892 (15.0)	.82
Currently doing sports[f]	342 (79.5)	265 (81.9)	.46	286 (78.0)	.63	291 (77.0)	.44	295 (80.3)	888 (78.6)	.52	295 (79.3)	888 (78.9)	.89
Media consumption													
Social media (>2 hours/day)[f]	341 (20.2)	262 (19.1)	.72	289 (17.3)	.35	290 (19.7)	.86	295 (20.0)	886 (18.8)	.67	295 (21.3)	886 (18.4)	.30
TV (>2 hours/day)[f]	341 (42.2)	262 (41.6)	.88	289 (41.2)	.79	292 (38.0)	.28	294 (39.1)	889 (41.4)	.49	294 (41.5)	889 (40.6)	.79
Game console (>2 hours/day)[f]	339 (18.9)	261 (19.9)	.75	287 (18.5)	.90	290 (12.8)	.03[i]	295 (20.0)	881 (16.7)	.21	295 (19.9)	881 (16.7)	.22

[a]Sample sizes are shown in Table 1.

[b]MM: mixed-mode.

[c]SAQ-Web: self-administered Web-based questionnaire.

[d]SAQ-paper: self-administered paper-based questionnaire.

[e]Adjusted for age of the child (adolescent, parent); sex of the child (adolescent); relationship to the child (parent); household income (adolescent, parent); parental education (adolescent, parent); adolescent education (adolescent); region (adolescent, parent); municipality size (adolescent, parent); and parental marital status (parent).

[f]Self-reported by adolescents aged 11-17 years.

[g]AUDIT-C: Alcohol Use Disorders Identification Test.

[h]WHO: World Health Organization; sample sizes are shown in Table 1.

[i]Not significant, *P* value adjusted with Bonferroni correction.

Discussion

Summary

The main aim of this study was to examine the risk of mode effects in a mixed-mode health interview survey for children and adolescents that combined paper-and-pencil questionnaires and Web-based questionnaires. Therefore, we compared prevalence rates and mean values of a broad range of health indicators from 3 alternative mixed-mode designs (all combining paper-and-pencil and Web-based questionnaires) with a single-mode control group (paper-and-pencil only). We also compared results between online respondents and paper-and-pencil respondents regardless of the survey design. First, we examined differences in sociodemographic characteristics by survey design and data collection mode, as it is well documented that sociodemographic characteristics are associated with health status and health behavior [43,44]. Regarding survey design, there were no statistically significant differences in sample composition, prevalence rates, or mean values of the examined health indicators. There were differences in sociodemographic characteristics across the data collection mode groups. After adjusting for these differences, only one of the analyzed health indicators (lifetime alcohol consumption) showed between-group differences. These results indicate that there is limited mode bias in health interview surveys for children and adolescents using a mixed-mode design with Web-based and paper questionnaires.

Sample Composition and Digital Divide

Consistent with previous findings, the sample composition of responding parents and of responding adolescents differed by data collection mode. We confirmed the so-called "digital divide" [47-50]—male adolescents and younger fathers preferred the online mode, a well-known systematic difference [5] between these modes[10,49,51-54]. Additionally, SAQ-Web respondents had higher household incomes [15,49,55] and higher household education levels [10,49,54-57]. Despite these differences, and differences in online response rates between the mixed-mode survey designs, there were no statistically significant differences in sample composition between the paper-and-pencil single-mode control group and the 3 mixed-mode groups. To control for the influence of sociodemographic on health indicators, we adjusted for sociodemographic characteristics by first calculating crude prevalence rates. Then, the analysis was complemented with adjusted prevalence rates or adjusted mean values using predictive margins to identify possible mode effects. Comparisons between the mixed-mode survey designs and the single-mode control group were made using only the crude prevalence rates. Using this approach, hardly any statistically significant differences by data collection mode or by survey design were found for the analyzed health indicators.

Health Status and Health Care Utilization

Prevalence rates of health complaints, such as diagnosed allergies, diagnosed ADHD, obesity, and chronic pain, were equivalent between the modes, as previous studies of adults [9,11,53,58] and adolescents [12] have shown. A population-based Norwegian study found higher asthma prevalence rates among online respondents; this was interpreted as possible nonresponse bias and not as a mode effect because there were no differences in the prevalence rates for any other condition [59]. A literature review by Hox et al showed that after controlling for selection, small mode effects do appear, most often distinguishing between modes that involve interviewers (face-to-face, telephone) and modes that do not (mail, Web) [42].

We found similar prevalence rates for SRH, chronic diseases, and impairment owing to health problems between SAQ-paper and SAQ-Web respondents. The 2 previous studies examining these health indicators among adults in general [11] and among older adults [10] also found no differences between these 2 data collection modes. Another study of adults interpreted the higher SRH found among online respondents compared with paper-based respondents as an expression of different sample characteristics linked to the digital divide era [49], or a case of better-situated people with better health using Web-based questionnaires, and not as a mode effect. We cannot say whether this holds true for the KiGGS methodological pilot study, because we controlled for most characteristics linked to the preference for online participation, such as region of residence and education or income.

For mental and psychosocial problems, we calculated risk groups for emotional and behavioral problems and for impairment owing to psychosocial problems based on SDQ scores [34]. Both parent- and adolescent-reported scores were equivalent across the examined modes. Several other studies have postulated the comparability of measurement results between these 2 self-administered modes for other standardized mental health questionnaires (eg, depression or anxiety) [12,58,60,61].

In their review of 55 studies investigating 79 instruments, Campbell et al [9] found measurement equivalence for electronic- and paper-based patient outcomes and concluded that standardized instruments can generally be used electronically without measurement effects. In our study, we also found comparable results for standardized instruments (the SDQ and AUDIT-C), as well as for self-reported HRQoL (KIDSCREEN-27). No existing studies have compared these particular instruments, but previous studies have compared the Short Form Health Survey-36, a frequently used standardized HRQoL instrument for adults, and found measurement equivalence [9,58,62-64].

All reports of health care utilization were equivalent between the self-administered modes; this is consistent with prior empirical results, including studies of adult vaccination use [11], adolescent health care use [12], and multiple health care quality indicators [56]. The greater use of pediatric services (and of any doctor) before adjustment for sociodemographic characteristics may be explained by the younger age of children in the online group—in Germany, all children are invited to undergo regular health screening examinations (U3-U9 examinations) from early childhood until the age of 5 years, with a well-established system of reminders and reporting.

Health Behaviors

Most of the analyzed adolescent health behaviors (current smoking, 12-month water pipe consumption, second-hand smoke exposure, physical activity, and screen-based media use) showed comparable results and no differences between the 2 modes. These results are consistent with the results of other studies on adolescents [12,65].

Considering alcohol consumption, the crude and adjusted prevalence rates for lifetime consumption were significantly higher among SAQ-Web-responding adolescents. After adjusting for sociodemographic characteristics, the difference decreased but could not be explained by the sociodemographic differences between the 2 groups of respondents. The prevalence of hazardous consumption and binge drinking were comparable between data collection modes after controlling for sample composition.

Most previous studies have reported no statistically significant differences in alcohol consumption among adolescents or young adults by these 2 data collection modes [12,66]. However, research comparing sensitive health behaviors is inconsistent. Some studies have found higher adult binge drinking [53] and higher adolescent alcohol consumption [17] in online reports, whereas others have found no difference in sensitive health behaviors in general for college students [67,68] and young adults [69].

The higher rate of reported lifetime alcohol consumption among SAQ-Web-responding adolescents, in the absence of frequently reported hazardous consumption or binge drinking, may be interpreted in multiple ways. For example, this may be a result of different sample properties, such as SAQ-Web-preferring adolescents being more likely to experiment with alcohol consumption. However, it is also possible that this result is a mode effect based on the assumption of identical alcohol consumption in both groups. Web-based questionnaires afford greater privacy because there is no risk of parents checking the responses. Another possible explanation is the lower social orientation in the Internet mode [13]. Both these explanations assume that Web-based questionnaires are more likely to elicit honest reports, but the similar results between the 2 mode groups for reported harmful alcohol consumption after adjustment contradict this assumption. Taken together, the results for alcohol consumption suggest that lifetime consumption should be used with caution as a health indicator in a mixed-mode design. Hazardous consumption and binge drinking are better indicators because they exhibit mode equivalence and have greater public health relevance than lifetime consumption, which is measured by a single question asking whether the respondent has ever consumed alcohol.

Main Result

Other empirical comparisons of measurement results between different mixed-mode survey designs are rare. In accord with one other result for the adult population [70], all of the analyzed health indicators for children and adolescents showed comparable results, with no statistically significant differences between the single-mode control group and the 3 mixed-mode groups. Additionally, sociodemographic characteristics did not differ by survey design for parents or adolescents. Regarding measurement comparability, any of the tested mixed-mode health interview survey designs, which offer both Web-based and paper questionnaires, could be used for children and adolescents.

Strengths and Limitations

The strengths of the methodological pilot study are the randomized study design, the population-based sample, and the inclusion of a single-mode control group as a reference to interpret the results. However, there are also some limitations, predominantly the relatively small size of the net samples of the analyzed groups. Each survey design had a relatively low number of cases, so interpretations of the results based on the net samples must be made with caution. Possible differences across the 4 survey designs or between the 2 data collection modes could have been overlooked because of a lack of statistical power, particularly regarding the need for correction for multiple testing. Other limitations concern the external validity of the results; the study was conducted in a German setting using register-based samples of children and adolescents, so the results are difficult to generalize to other countries, settings, or populations.

Conclusions

Our results are consistent with those of most previous studies. We found comparable results between the 2 self-administered modes (SAQ-Web and SAQ-paper) for almost all analyzed health indicators, except for lifetime consumption of alcohol among adolescents aged 11-17 years. Thus, no differences were found between the single-mode control group design and 3 mixed-mode survey designs that combined the 2 data collection modes.

These results suggest that it is possible to measure health indicators for children and adolescents using a mixed-mode design combining SAQ-Web and SAQ-paper methods, with a low risk of mode effects and high comparability across different mixed-mode survey designs combining these 2 data collection modes [4]. The implementation of a Web-based option in the existing paper-based interview surveys of children and adolescents has a low risk of changed measurement values caused by the mixed-mode survey design.

Acknowledgments

This study was funded by the Robert Koch Institute and the German Federal Ministry of Health within the German Health Monitoring System. The authors thank Diane Williams, PhD, from Edanz Group for editing a draft of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Checklist for Reporting Results of Internet E-Surveys (CHERRIES).

[PDF File (Adobe PDF File), 92KB - jmir_v20i3e64_app1.pdf]

Multimedia Appendix 2

Sociodemographic characteristics of responding parents of children aged 0-17 years by survey design and data collection mode mode.

[PDF File (Adobe PDF File), 58KB - jmir_v20i3e64_app2.pdf]

Multimedia Appendix 3

Sociodemographic characteristics of responding adolescents aged 11-17 years by survey design and data collection mode.

[PDF File (Adobe PDF File), 50KB - jmir_v20i3e64_app3.pdf]

References

1. Galea S, Tracy M. Participation rates in epidemiologic studies. Ann Epidemiol 2007 Sep;17(9):643-653. [doi: 10.1016/j.annepidem.2007.03.013] [Medline: 17553702]
2. Bowling A. Mode of questionnaire administration can have serious effects on data quality. J Public Health (Oxf) 2005 Sep;27(3):281-291. [doi: 10.1093/pubmed/fdi031] [Medline: 15870099]
3. Revilla M. Quality in unimode and mixed-mode designs: a multitrait-multimethod approach. Surv Res Methods 2010;4(3):151-164 [FREE Full text] [doi: 10.18148/srm/2010.v4i3.4278]
4. De Leeuw E, Hox J. Internet surveys as part of a mixed mode design. In: Das M, Ester P, Kaczmirek L, editors. Social and Behavioral Research and the Internet: Advances in Applied Methods and Research Strategies. New York: Taylor & Francis Group; 2011:45-76.
5. Hox JJ, De Leeuw ED, Zijlmans EA. Measurement equivalence in mixed mode surveys. Front Psychol 2015;6:87 [FREE Full text] [doi: 10.3389/fpsyg.2015.00087] [Medline: 25699002]
6. Vannieuwenhuyze J, Loosveldt G, Molenberghs G. A method for evaluating mode effects in mixed-mode surveys. Public Opin Q 2011 Feb 09;74(5):1027-1045. [doi: 10.1093/poq/nfq059]
7. Couper MP. The future of modes of data collection. Public Opin Q 2011 Dec 15;75(5):889-908. [doi: 10.1093/poq/nfr046]
8. Rutherford C, Costa D, Mercieca-Bebber R, Rice H, Gabb L, King M. Mode of administration does not cause bias in patient-reported outcome results: a meta-analysis. Qual Life Res 2016 Mar 3;25(3):559-574. [doi: 10.1007/s11136-015-1110-8] [Medline: 26334842]
9. Campbell N, Ali F, Finlay AY, Salek SS. Equivalence of electronic and paper-based patient-reported outcome measures. Qual Life Res 2015 Aug;24(8):1949-1961. [doi: 10.1007/s11136-015-0937-3] [Medline: 25702266]
10. de Bernardo DH, Curtis A. Using online and paper surveys: the effectiveness of mixed-mode methodology for populations over 50. Res Aging 2013;35(2):220-240. [doi: 10.1177/0164027512441611]
11. Hoebel J, von der Lippe E, Lange C, Ziese T. Mode differences in a mixed-mode health interview survey among adults. Arch Public Health 2014 Dec 22;72(1):46 [FREE Full text] [doi: 10.1186/2049-3258-72-46] [Medline: 25810913]
12. Mangunkusumo RT, Moorman PW, Van Den Berg-de Ruiter AE, Van Der Lei J, De Koning HJ, Raat H. Internet-administered adolescent health questionnaires compared with a paper version in a randomized study. J Adolesc Health 2005 Jan;36(1):70.e1-70.e6. [doi: 10.1016/j.jadohealth.2004.02.020]
13. Taddicken M. Methodeneffekte von Web-Befragungen: Soziale Erwünschtheit vs. Soziale Entkontextualisierung. In: Weichbold M, Bacher J, Wolf C, editors. Umfrageforschung: Herausforderungen und Grenzen. Wiesbaden: VS Verlag für Sozialwissenschaften; 2009:85-104.
14. van Gelder MM, Bretveld RW, Roeleveld N. Web-based questionnaires: the future in epidemiology? Am J Epidemiol 2010 Dec 01;172(11):1292-1298. [doi: 10.1093/aje/kwq291] [Medline: 20880962]
15. Rookey BD, Hanway S, Dillman DA. Does a probability-based household panel benefit from assignment to postal response as an alternative to Internet-only? Public Opin Q 2008 Dec 09;72(5):962-984. [doi: 10.1093/poq/nfn061]
16. Burkill S, Copas A, Couper MP, Clifton S, Prah P, Datta J, et al. Using the web to collect data on sensitive behaviours: a study looking at mode effects on the British national survey of sexual attitudes and lifestyles. PLoS One 2016 Feb 11;11(2):e0147983 [FREE Full text] [doi: 10.1371/journal.pone.0147983] [Medline: 26866687]
17. Paperny DM, Aono JY, Lehman RM, Hammar SL, Risser J. Computer-assisted detection and intervention in adolescent high-risk health behaviors. J Pediatr 1990 Mar;116(3):456-462. [Medline: 2308041]

18. Kurth BM, Kamtsiuris P, Hölling H, Schlaud M, Dölle R, Ellert U, et al. The challenge of comprehensively mapping children's health in a nation-wide health survey: design of the German KiGGS-Study. BMC Public Health 2008 Jun 04;8:196 [FREE Full text] [doi: 10.1186/1471-2458-8-196] [Medline: 18533019]
19. Hölling H, Schlack R, Kamtsiuris P, Butschalowsky H, Schlaud M, Kurth BM. [The KiGGS study. Nationwide representative longitudinal and cross-sectional study on the health of children and adolescents within the framework of health monitoring at the Robert Koch Institute]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2012 Jun 7;55(6-7):836-842. [doi: 10.1007/s00103-012-1486-3] [Medline: 22736165]
20. Kurth BM. [Das RKI-Gesundheitsmonitoring - was es enthält und wie es genutzt werden kann]. Public Health Forum 2012;20(3):4.e1-4.e3. [doi: 10.1016/j.phf.2012.06.001]
21. Kurth BM, Kamtsiuris P, Hölling H, Mauz E. Strategien des Robert Koch-Instituts zum Monitoring der Gesundheit von in Deutschland lebenden Kindern und Jugendlichen. Kinder- und Jugendmedizin 2016;16(3):176-183. [doi: 10.1055/s-0037-1616317]
22. Robert Koch-Institut. Rki.de. 2011. KiGGS - Kinder- und Jugendgesundheitsstudie Welle 1. Projektbeschreibung URL: https://www.rki.de/DE/Content/Gesundheitsmonitoring/Gesundheitsberichterstattung/GBEDownloadsB/KiGGS_welle1.pdf?__blob=publicationFile [accessed 2018-01-30] [WebCite Cache ID 6ws4yLl7h]
23. Mauz E, Gößwald A, Kamtsiuris P, Hoffmann R, Lange M, von Schenck U, et al. New data for action. Data collection for KiGGS Wave 2 has been completed. J Health Monit 2017;2(3):2-27 [FREE Full text] [doi: 10.17886/RKI-GBE-2017-105]
24. Biemer PP. Total survey error: design, implementation, and evaluation. Public Opin Q 2011 Feb 09;74(5):817-848. [doi: 10.1093/poq/nfq058]
25. Groves RM, Lyberg L. Total survey error: past, present, and future. Public Opin Q 2011 Feb 09;74(5):849-879. [doi: 10.1093/poq/nfq065]
26. Robert Koch-Institut. Berlin: Robert Koch-Institut; 2017. Methodische Studie zur Durchführung von Mixed-Mode-Befragungen zur Gesundheit von Kindern und Jugendlichen (Pilotstudie KiGGS Welle 2) URL: https://www.rki.de/DE/Content/Gesundheitsmonitoring/Gesundheitsberichterstattung/GBEDownloadsB/Pilotstudie-KiGGS-Welle2.pdf?__blob=publicationFile [accessed 2018-01-27] [WebCite Cache ID 6wml7CUYs]
27. Dillman DA, Smyth JD, Christian LM. Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method. 4th Edition. New York: John Wiley & Sons; 2014.
28. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Dec 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
29. American Association for Public Opinion Research (AAPOR). Esomar. 2011. Standard definitions: final dispositions of case codes and outcome rates for surveys (revised 2011) URL: https://www.esomar.org/uploads/public/knowledge-and-standards/codes-and-guidelines/ESOMAR_Standard-Definitions-Final-Dispositions-of-Case-Codes-and-Outcome-Rates-for-Surveys.pdf [accessed 2018-01-30] [WebCite Cache ID 6wrLvtFT3]
30. Brauns H, Scherer S, Steinmann S. The CASMIN Educational Classification in International Comparative Research. In: Hoffmeyer-Zlotnik JHP, Wolf C. editors. Advances in Cross-National Comparison: A European Working Book for Demographic and Socio-Economic Variables. Boston, MA: Springer US; 2003:221-244.
31. Cox B, van Oyen H, Cambois E, Jagger C, le Roy S, Robine JM, et al. The reliability of the Minimum European Health Module. Int J Public Health 2009 Jan;54(2):55-60. [doi: 10.1007/s00038-009-7104-y] [Medline: 19183846]
32. Bethell CD, Read D, Stein RE, Blumberg SJ, Wells N, Newacheck PW. Identifying children with special health care needs: development and evaluation of a short screening instrument. Ambul Pediatr 2002;2(1):38-48. [Medline: 11888437]
33. Kromeyer-Hauschild K, Wabitsch M, Kunze D, Geller F, Geiß H, Hesse V, et al. Perzentile für den Body-mass-Index für das Kindes- und Jugendalter unter Heranziehung verschiedener deutscher Stichproben. Monatsschr Kinderheilkd 2001 Aug 1;149(8):807-818 [FREE Full text] [doi: 10.1007/s001120170107]
34. Goodman R. The extended version of the Strengths and Difficulties Questionnaire as a guide to child psychiatric caseness and consequent burden. J Child Psychol Psychiatry 1999 Jul;40(5):791-799. [Medline: 10433412]
35. Woerner W, Becker A, Rothenberger A. Normative data and scale properties of the German parent SDQ. Eur Child Adolesc Psychiatry 2004 Jul;13 Suppl 2:II3-I10. [doi: 10.1007/s00787-004-2002-6] [Medline: 15243780]
36. Rattay P, Starker A, Domanska O, Butschalowsky H, Gutsche J, Kamtsiuris P, KiGGS Study Group. [Trends in the utilization of outpatient medical care in childhood and adolescence: results of the KiGGS study - a comparison of baseline and first follow up (KiGGS Wave 1)]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2014 Jul;57(7):878-891. [doi: 10.1007/s00103-014-1989-1] [Medline: 24950837]
37. The Kidscreen Group Europe (editors). The KIDSCREEN Questionnaires: Quality of Life Questionnaires for Children and Adolescents. Lengerich, Germany: Pabst Science Publishers; 2006.
38. Lampert T, Kuntz B, KiGGS Study Group. [Tobacco and alcohol consumption among 11- to 17-year-old adolescents: results of the KiGGS study: first follow-up (KiGGS Wave 1)]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2014 Jul;57(7):830-839. [doi: 10.1007/s00103-014-1982-8] [Medline: 24950832]
39. Manz K, Schlack R, Poethko-Müller C, Mensink G, Finger J, Lampert T, KiGGS Study Group. [Physical activity and electronic media use in children and adolescents: results of the KiGGS study: first follow-up (KiGGS wave 1)].

Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2014 Jul;57(7):840-848. [doi: 10.1007/s00103-014-1986-4] [Medline: 24950833]
40. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test. Arch Intern Med 1998 Sep 14;158(16):1789-1795. [Medline: 9738608]
41. World Health Organization. Geneve: WHO; 2010. Global recommendations on physical activity for health URL: http://www.who.int/dietphysicalactivity/publications/9789241599979/en/ [accessed 2018-01-27] [WebCite Cache ID 6wmgoTDWv]
42. Hox J, De LE, Klausch T. Mixed mode research: issues in design and analysis. In: Biemer PP, de Leeuw E, Eckman S, Edwards B, Kreuter F, Lyberg LE, et al, editors. Total Survey Error in Practice: Improving Quality in the Era of Big Data. New Yorck: Wiley; 2017:511-530.
43. Abuladze L, Kunder N, Lang K, Vaask S. Associations between self-rated health and health behaviour among older adults in Estonia: a cross-sectional analysis. Br Med J Open 2017 Jun 09;7(6):e013257 [FREE Full text] [doi: 10.1136/bmjopen-2016-013257] [Medline: 28601816]
44. Link BG, Phelan JC. Understanding sociodemographic differences in health--the role of fundamental social causes. Am J Public Health 1996 Apr;86(4):471-473. [Medline: 8604773]
45. Graubard BI, Korn EL. Predictive margins with survey data. Biometrics 1999 Jun;55(2):652-659. [Medline: 11318229]
46. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. Br Med J 1995 Jan 21;310(6973):170 [FREE Full text] [Medline: 7833759]
47. Estacio EV, Whittle R, Protheroe J. The digital divide: examining socio-demographic factors associated with health literacy, access and use of internet to seek health information. J Health Psychol 2017 Feb 01:1359105317695429. [doi: 10.1177/1359105317695429] [Medline: 28810415]
48. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014 Jul;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
49. Shim JM, Shin E, Johnson TP. Self-rated health assessed by web versus mail modes in a mixed mode survey: the digital divide effect and the genuine survey mode effect. Med Care 2013 Sep;51(9):774-781. [doi: 10.1097/MLR.0b013e31829a4f92] [Medline: 23774510]
50. Latulippe K, Hamel C, Giroux D. Social health inequalities and eHealth: a literature review with qualitative synthesis of theoretical and empirical studies. J Med Internet Res 2017 Apr 27;19(4):e136 [FREE Full text] [doi: 10.2196/jmir.6731] [Medline: 28450271]
51. Börkan B. The mode effect in mixed-mode surveys mail and web surveys. SSCR 2010;28(3):371-380 [FREE Full text]
52. Dolnicar S, Laesser C, Matus K. Online versus paper format effects in tourism surveys. J Travel Res 2009;47(3):295-316.
53. Link MW, Mokdad AH. Alternative modes for health surveillance surveys: an experiment with web, mail, and telephone. Epidemiology 2005 Sep;16(5):701-704. [Medline: 16135951]
54. Callas PW, Solomon LJ, Hughes JR, Livingston AE. The influence of response mode on study results: offering cigarette smokers a choice of postal or online completion of a survey. J Med Internet Res 2010 Oct 21;12(4):e46 [FREE Full text] [doi: 10.2196/jmir.1414] [Medline: 20965873]
55. Messer B, Edwards M, Dillman D. Determinants of item nonresponse to web and mail respondents in three address-based mixed-mode surveys of the general public. Surv Pract 2012;5(2):1-27 [FREE Full text]
56. 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 Sep 27;13(3):e68 [FREE Full text] [doi: 10.2196/jmir.1241] [Medline: 21946048]
57. Smith AB, King M, Butow P, Olver I. A comparison of data quality and practicality of online versus postal questionnaires in a sample of testicular cancer survivors. Psychooncology 2013 Jan;22(1):233-237. [doi: 10.1002/pon.2052] [Medline: 21954102]
58. Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. Value Health 2008;11(2):322-333 [FREE Full text] [doi: 10.1111/j.1524-4733.2007.00231.x] [Medline: 18380645]
59. Brøgger J, Nystad W, Cappelen I, Bakke P. No increase in response rate by adding a web response option to a postal population survey: a randomized trial. J Med Internet Res 2007 Dec 31;9(5):e40 [FREE Full text] [doi: 10.2196/jmir.9.5.e40] [Medline: 18174120]
60. Austin DW, Carlbring P, Richards JC, Andersson G. Internet administration of three commonly used questionnaires in panic research: equivalence to paper administration in Australian and Swedish samples of people with panic disorder. Int J Test 2006 Mar;6(1):25-39. [doi: 10.1207/s15327574ijt0601_2]
61. Zhang X, Kuchinke L, Woud ML, Velten J, Margraf J. Survey method matters: online/offline questionnaires and face-to-face or telephone interviews differ. Comput Human Behav 2017 Jun;71:172-180. [doi: 10.1016/j.chb.2017.02.006]
62. 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]

63. Broering JM, Paciorek A, Carroll PR, Wilson LS, Litwin MS, Miaskowski C. Measurement equivalence using a mixed-mode approach to administer health-related quality of life instruments. Qual Life Res 2014 Mar;23(2):495-508. [doi: 10.1007/s11136-013-0493-7] [Medline: 23943258]
64. 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]
65. Wyrick DL, Bond L. Reducing sensitive survey response bias in research on adolescents: a comparison of web-based and paper-and-pencil administration. Am J Health Promot 2011 May;25(5):349-352. [doi: 10.4278/ajhp.080611-ARB-90] [Medline: 21534838]
66. McCabe SE, Diez A, Boyd CJ, Nelson TF, Weitzman ER. Comparing web and mail responses in a mixed mode survey in college alcohol use research. Addict Behav 2006 Sep;31(9):1619-1627 [FREE Full text] [doi: 10.1016/j.addbeh.2005.12.009] [Medline: 16460882]
67. Pealer LN, Weiler RM, Pigg RM, Miller D, Dorman SM. The feasibility of a web-based surveillance system to collect health risk behavior data from college students. Health Educ Behav 2001 Oct;28(5):547-559. [doi: 10.1177/109019810102800503] [Medline: 11575685]
68. Jun MK. Websm. Indiana: Indiana University; 2005. Effects of survey mode, gender, and perceived sensitivity on the quality of data regarding sensitive health behaviors URL: http://www.websm.org/db/12/11689/rec/ [accessed 2018-01-27] [WebCite Cache ID 6wmoeta7p]
69. McMorris BJ, Petrie RS, Catalano RF, Fleming CB, Haggerty KP, Abbott RD. Use of web and in-person survey modes to gather data from young adults on sex and drug use: an evaluation of cost, time, and survey error based on a randomized mixed-mode design. Eval Rev 2009 Apr;33(2):138-158 [FREE Full text] [doi: 10.1177/0193841X08326463] [Medline: 19029360]
70. Schilling R, Hoebel J, Müters S, Lange C. Edoc.rki.de. Berlin: RKI; 2015. Pilotstudie zur Durchführung von Mixed-Mode-Gesundheitsbefragungen in der Erwachsenenbevölkerung (Projektstudie GEDA 2.0). Beiträge zur Gesundheitsberichterstattung des Bundes URL: http://edoc.rki.de/documents/rki_fv/rex2KXJHIPpM/PDF/226m8YvEr7dPM.pdf [accessed 2018-01-27] [WebCite Cache ID 6wmkxBuML]

Abbreviations

ADHD: attention-deficit hyperactivity disorder
AUDIT-C: Alcohol Use Disorders Identification Test
HRQoL: health-related quality of life
KiGGS: The German Health Interview and Examination Survey for Children and Adolescents
MM: mixed-mode
SAQ-paper: self-administered paper-based questionnaires
SAQ-Web: self-administered Web-based questionnaires
SDQ: Strengths and Difficulties Questionnaire
SRH: self-rated health

Edited by G Eysenbach; submitted 12.04.17; peer-reviewed by T Johnson, J Broering; comments to author 15.06.17; revised version received 14.10.17; accepted 16.11.17; published 05.03.18.

Please cite as:
Mauz E, Hoffmann R, Houben R, Krause L, Kamtsiuris P, Gößwald A
Mode Equivalence of Health Indicators Between Data Collection Modes and Mixed-Mode Survey Designs in Population-Based Health Interview Surveys for Children and Adolescents: Methodological Study
J Med Internet Res 2018;20(3):e64
URL: http://www.jmir.org/2018/3/e64/
doi:10.2196/jmir.7802
PMID:29506967

Original Paper

Technology and Social Media Use Among Patients Enrolled in Outpatient Addiction Treatment Programs: Cross-Sectional Survey Study

Robert D Ashford[1], MSW; Kevin Lynch[1], PhD; Brenda Curtis[1], PhD

Addictions, Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, United States

Corresponding Author:
Robert D Ashford, MSW
Addictions
Department of Psychiatry
University of Pennsylvania
3440 Market Street, Suite 370
Philadelphia, PA, 19104
United States
Phone: 1 817 614 4302
Email: roberdav@upenn.edu

Abstract

Background: Substance use disorder research and practice have not yet taken advantage of emerging changes in communication patterns. While internet and social media use is widespread in the general population, little is known about how these mediums are used in substance use disorder treatment.

Objective: The aims of this paper were to provide data on patients' with substance use disorders mobile phone ownership rates, usage patterns on multiple digital platforms (social media, internet, computer, and mobile apps), and their interest in the use of these platforms to monitor personal recovery.

Methods: We conducted a cross-sectional survey of patients in 4 intensive outpatient substance use disorder treatment facilities in Philadelphia, PA, USA. Logistic regressions were used to examine associations among variables.

Results: Survey participants (N=259) were mostly male (72.9%, 188/259), African American (62.9%, 163/259), with annual incomes less than US $10,000 (62.5%, 161/259), and averaged 39 (SD 12.24) years of age. The vast majority of participants (93.8%, 243/259) owned a mobile phone and about 64.1% (166/259) owned a mobile phone with app capabilities, of which 85.1% (207/243) accessed the internet mainly through their mobile phone. There were no significant differences in age, gender, ethnicity, or socio-economic status by computer usage, internet usage, number of times participants changed their phone, type of mobile phone contract, or whether participants had unlimited calling plans. The sample was grouped into 3 age groups (Millennials, Generation Xers, and Baby Boomers). The rates of having a social media account differed across these 3 age groups with significant differences between Baby Boomers and both Generation Xers and Millennials (P<.001 in each case). Among participants with a social media account (73.6%, 190/259), most (76.1%, 144/190) reported using it daily and nearly all (98.2%, 186/190) used Facebook. Nearly half of participants (47.4%, 90/190) reported viewing content on social media that triggered substance cravings and an equal percentage reported being exposed to recovery information on social media. There was a significant difference in rates of reporting viewing recovery information on social media across the 3 age groups with Baby Boomers reporting higher rates than Millennials (P<.001). The majority of respondents (70.1%, 181/259) said they would prefer to use a relapse prevention app on their phone or receive SMS (short message service) relapse prevention text messages (72.3%, 186/259), and nearly half (49.1%, 127/259) expressed an interest in receiving support by allowing social media accounts to be monitored as a relapse prevention technique.

Conclusions: To our knowledge, this is the first and largest study examining the online behavior and preferences regarding technology-based substance use disorder treatment interventions in a population of patients enrolled in community outpatient treatment programs. Patients were generally receptive to using relapse prevention apps and text messaging interventions and a substantial proportion supported social media surveillance tools. However, the design of technology-based interventions remains as many participants have monthly telephone plans which may limit continuity.

(J Med Internet Res 2018;20(3):e84) doi:10.2196/jmir.9172

KEYWORDS

digital divide; behavioral health; social media; addiction; recovery; relapse

Introduction

Mobile phone use has increased dramatically over the last decade. Today, 95% of US adults have a mobile phone, up from 66% in 2006. Mobile phone with app capability (smartphone) ownership has also more than doubled from 35% in 2011 to 77% in 2016 [1]. With the increase in mobile phone ownership, health care providers have become interested in integrating the use of mobile phones in the care of chronic conditions such as HIV, diabetes, hypertension, and asthma [2-6]. Mobile phones are portable, capable of receiving and transmitting data, and they are "always on." They also offer health care providers the unique ability to connect with hard-to-reach populations that might otherwise not have access to health care services [7,8]. Mobile phones with app capabilities have the added benefit of being mini-computers that run software apps, connect to the internet, and have various embedded sensors. Sensors include Global Positioning Systems that monitor spatial location, accelerometers that record movement and gross motor activities, text that can be automatically analyzed using natural language processing to provide users with personalized feedback, and voice and tone records that can estimate mood [9]. Mobile phones also provide health care providers with a continuous stream of patient data regarding behavior, symptoms, and physiology. Recently, behavioral health apps have been developed that monitor psychiatric illnesses [10-14]. These apps have proven to be feasible across a wide range of conditions including schizophrenia [15], bipolar disorder [16], anxiety disorder [17], and depression [14,18,19].

Addiction researchers have recently begun to explore the use of mobile phones to support recovery from substance use disorders [20]. Alcohol and drug misuse was estimated to cost the nation over US $400 billion annually. In 2015, an estimated 7.7 million individuals in the United States had an illicit drug use disorder in the past year, but only 1.3 million individuals (approximately 17%) received substance use treatment [21]. Further, fewer than half of the patients who entered treatment completed it [22], with about 70% of patients experiencing a recurrence of use within a few months of initiating treatment [23].

Mobile phones can provide access to an assortment of online resources for attendees of outpatient substance use disorder treatment programs [24]. For example, patients can access the internet to search for local Alcoholics Anonymous (AA) and/or Narcotics Anonymous (NA) meetings, attend online AA/NA meetings, join an online recovery community, or download a recovery-based app to their mobile phone. Patients in treatment may also be able to stay connected to loved ones via phone and video calls, texting, and social networking sites, providing them with needed social support. These individuals can also use their phones to make medical appointments, communicate with health care providers, and communicate with potential employers.

Little is known about access to mobile phones and the use of digital platforms among patients attending outpatient substance use disorder treatment programs. McClure and colleagues examined the utilization of the internet and mobile phones with patients attending outpatient clinics, methadone/buprenorphine maintenance programs, and buprenorphine maintenance primary care clinics [25]. The majority of the patients reported access to a mobile phone (91%) and texting (79%). Patients also reported higher regular internet use (44%) than regular computer use (39%), which suggested that some patients were accessing the internet through their mobile phones, though information on the type of phone used was not collected in the study.

This paper explored the patterns of mobile phone usage, the use of Facebook, Twitter, and other social media platforms, and the receptiveness of using these types of digital platforms for interventions that promote positive recovery outcomes among patients attending community outpatient substance use disorder treatment programs. We also examined differences in digital platform use by age, gender, income, race, and preferred substances. The goals of these analyses were to assess disparities in digital media platform use and gain a better understanding of which platforms would be suitable for disseminating and sustaining real-world recovery promotion interventions for this population.

Methods

Recruitment

In May 2016, a self-administered, in-person, paper-and-pencil survey was conducted among patients attending outpatient substance use disorder treatment programs in Philadelphia. Participants were invited to participate by research staff between group sessions. These outpatient treatment programs treated approximately 800 patients monthly, all over 18 years of age. The requirements to participate in the survey were (1) current enrollment in the outpatient program at the time of the survey; (2) greater than 18 years of age; (3) no intellectual or developmental disability; (4) and willingness to provide informed consent to participate. The survey took 10 minutes to complete and no identifiable information was recorded to protect patient privacy. All study procedures were approved by the University of Pennsylvania Human Subjects Review Board.

Survey

The survey included technology utilization questions, adapted from McClure and colleagues [25], to assess the communication patterns of patients enrolled in substance abuse treatment programs in the Baltimore area. The survey was updated to include questions pertaining to mobile phone ownership, social media usage, and interest in the use of digital platforms to monitor recovery (Multimedia Appendix 1).

Exposure level to drug cues and pro-recovery information on social media was measured via responses to the following items:

How often have you seen drug cues—things that made you want to use drugs on social media? [Responses ranging from 1 (always) to 5 (never)]

How often have you seen recovery information on social media? [Responses ranging from 1 (always) to 5 (never)]

Have you posted information on social media about being in recovery? [Yes/no]

We measured receptiveness to the use of online platforms for interventions that promote positive substance use treatment outcomes via responses to the following items:

Do you think social media would be a good place to receive information to help you avoid relapse? [Yes/no]

Would you join an online support group to help you during your recovery? [Yes/no]

Would you join a Facebook support group to help you during your recovery? [Yes/no]

Would you sign up to receive text messages to help you during your recovery? [Yes/no]

Would you use an app placed on your phone to help your recovery from alcohol or substance use? [Yes/no]

We also asked participants how they would like to access a digital outpatient treatment program to aide during recovery (website, social media, texting, app), and if they would you allow their social media accounts to be monitored to help prevent relapse.

Data Analysis

The responses were entered into the data monitoring system using double entry. One research assistant entered the data while checking for mismatches and out-of-range values. A different research assistant then entered the same data again. The entries were compared via a computer that identified mismatches. When mismatches were identified, the data entry persons checked the original survey to determine the correct value(s).

All analyses were performed using SAS 9.3. The primary comparisons were of technology ownership and use responses over 3 age groups (Baby Boomers, Generation Xers, and Millennials), using binary and ordinal logistic regression models. These models included gender and race as covariates.

Data Exclusion

Respondents were excluded from final analyses due to missing age, gender, race, and reporting zero use of technology. Respondents excluded due to missing demographic information totaled 6 (0.02%, 6/276) participants. Respondents excluded due to reporting zero use of technology totaled 11 (0.04%, 11/276) participants.

Results

Participants

Demographic information for study participants is shown in Table 1. The participants were 259 adults with a substance use disorder at the time of entrance to treatment at 4 Philadelphia area community-based intensive outpatient programs. The participants averaged 38.86 (SD 12.24) years of, primarily self-identified as male (72.9%, 188/259), African American (62.9%, 163/259), high school graduate or General Equivalency Diploma (GED; 58.8%, 152/259), single-never married (73.5%, 190/259), and unemployed (77.6%, 201/259) with a yearly income under US $10,000 (62.4%, 161/259). In addition, participants were stratified into the Millennial (18 to 35 years), Generation X (36 to 51 years), or Baby Boomer or older (52 or more years) generational categories that best mirror generational categories among the general population. Most participants were of the Millennial generation (46.3%, 120/259), followed by Generation X (32.4%, 84/259), and Baby Boomer or older (21.2%, 55/259). Participants cited marijuana (47.9%, 124/259) and alcohol (40.5%, 105/259) as preferred substances most frequently and the mean length of treatment was 4.64 (SD 8.35) months.

Technology Ownership

Of the 259 participants, most owned a mobile phone (93.8%, 243/259), of which many were mobile phones with app capabilities (64.1%, 166/259), with no significant differences between generations (χ^2_2=1.39; *P*=.50). Among phone owners, mobile phone with app capability ownership differed significantly among generations (χ^2_2=17.62, *P*<.001); with Generation X (*P*=.001, OR 3.52 [95% CI 1.65-7.52]) and Millennial (*P*<. 001, OR 4.53 [95% CI 2.19-9.35]) generations being more likely to own a mobile phone with capabilities than Baby Boomers. No significant differences were found between Millennials and Generation Xers (*P*=.11, OR 0.78 [95% CI 0.40-1.51]). Among phone owners, provider plans differed significantly among generations (χ^2_2=10.25, *P*=.006); with Generation Xers (*P*=.02, OR 3.73 [95% CI 1.18-11.73]) and Millennials (*P*=.002, OR 6.09 [95% CI 1.94-19.09]) being more likely to have unlimited texting plans than Baby Boomers.

Technology Use

The majority of all participants reported regularly using text messaging, email, the internet, and a computer. Generational differences were found to be significant in text message (χ^2_2=12.16, *P*=.002), email (χ^2_2=20.65, *P*<.001), and internet use (χ^2_2=26.37, *P*<.001), but not in computer use (χ^2_2=5.49, *P*=.06). The significant differences were largely due to the Baby Boomers using these media less than the Generation Xers or Millennials, with the Generation Xers tending to use less than the Millennials, but not significantly so (Multimedia Appendix 2).

Among those who accessed the internet, about 80.0% (152/190) reported that they typically accessed it via their mobile phone. There were no significant generational differences on accessing the internet by phone versus by some other means (χ^2_2=5.00, *P*=.08).

Social Media Ownership and Use

Of the respondents, 73.6% (190/259) had a social media account of some type, with the majority using these accounts daily

(76.1%, 144/190). There were significant generational differences (χ^2_2=38.25, *P*<.001); with both Generation Xers (*P*<.001, OR 5.62 [95% CI 2.62-12.03]) and Millennials (*P*<.001, OR 9.01 [95% CI 4.26-19.03]) being more likely to own a social media account. No differences were found between Millennials and Generation Xers (*P*=.19, OR 1.60 [95% CI 0.79-3.28]). There was a similar pattern of significant differences on frequency of use among participants with a social media account (χ^2_2=7.04, *P*=.03); with both Generation Xers (*P*=.02, OR 3.44 [95% CI 1.19-9.98]) and Millennials (*P*=.01, OR 3.81 [95% CI 1.38-10.59]) being more likely to have daily or weekly frequencies of use compared to Baby Boomers. No significant difference was found between Millennials and Generation Xers (*P*=.10, OR 1.11 [95% CI 0.51-2.41]).

Table 1. Demographic characteristics (N=259).

Characteristic	Value
Age in years, mean (SD)	38.86 (12.24)
Generation, n (%)	
Millennials	120 (46.3)
Generation X	84 (32.4)
Baby Boomer	55 (21.2)
Gender, n (%)	
Female	71 (27.0)
Male	188 (73.0)
Race, n (%)	
Nonblack	96 (37.1)
Black	163 (62.9)
Marital status, n (%)	
Single or never married	190 (73.5)
Married or domestic partnership	29 (11.3)
Widowed, divorced, or separated	40 (15.2)
Education level, n (%)	
Did not complete high school	77 (29.6)
High school graduate or GED[a]	152 (58.8)
2-year degree or more	30 (11.7)
Employment status, n (%)	
Employed	58 (22.4)
Unemployed	201 (77.6)
Income level per year, n (%)	
Less than $10,000	161 (62.5)
$10,000 to $49,999	87 (33.8)
Over $50,000	11 (3.8)
Substance use[b], n (%)	
Alcohol	105 (40.5)
Opiates	73 (28.2)
Cocaine	80 (30.9)
Amphetamines	17 (6.6)
Marijuana	124 (47.9)
Treatment length in months, mean (SD)	4.64 (8.35)

[a]GED: General Equivalency Diploma.

[b]Total is greater than 100% due to multiple responses from participants.

Virtually all participants with a social media account used Facebook (98.2%, 186/190), Instagram (60.22%, 114/190), Google+ (40.9%, 78/190), and Twitter (24.3%, 46/190) most frequently. Of these 4 main uses, there were significant generational differences for Twitter (χ^2_2=8.96, *P*=.03), Instagram (χ^2_2=24.68, *P*<.001), and SnapChat (χ^2_2=8.23, *P*=.02), but not for Google+ (χ^2_2=5.46, *P*=.07). Overall, Millennials made more use of all of the platforms than Baby Boomers or Generation Xers, with the exception of Google+.

Respondents predominantly used social media accounts to share photos and videos (83.4%, 158/190), stay in touch with family and friends (76.8%, 146/190), watch videos others post (70.2%, 133/190), instant message (67.4%, 128/190), and see updates about others (67.4%, 128/190).

Substance Use and Recovery on Social Media

Among people with social media accounts, 47.4% (90/190) of respondents had seen information (eg, posts, text, images, videos, etc) that made them want to use substances at least sometimes on digital media platforms, with exactly the same percentage reporting that they had seen recovery information at least sometimes. A cross-tabulation showed no significant association between the people in each of the reporting groups (χ^2_1=3.20, *P*=.07), and this was true within generation groups (*P*>.11 in each generation). There was no significant difference among generations on seeing drug cues (χ^2_2=3.14, *P*=.21). However, there was a significant generational difference (χ^2_2=8.39, *P*=.02) on the frequency of seeing recovery cues. The Millennials were significantly less likely to see recovery information than the Baby Boomers (*P*=.01, OR 0.31 [95% CI 0.12-0.77]), with no significant difference between Generation Xers and Baby Boomers (*P*=.19, OR 0.5 [95% CI 0.20-1.36]), or between Millennials and Generation Xers (*P*=.07, OR 0.58 [95% CI 0.32-1.05]).

The majority of respondents (59.7%, 113/190) had not posted about their personal recovery on social media accounts, and this did not significantly differ among generations (χ^2_2=4.53, *P*=.11).

Support on Social Media

Of the respondents, 66.1% (171/259) believed that social media platforms would be a good place to receive information to protect their recovery or prevent relapse. This belief did not differ across generations significantly (χ^2_2=6.11, *P*=.05). The majority of respondents (50.9%, 132/259) would not allow social media accounts to be monitored in order to support personal recovery, with no significant differences across generations (χ^2_2=2.34, *P*=.31).

Participants believed that providing support through social media (50.4%, 130/259) is preferred, compared to a website (36.8%, 95/259), text messaging (37.6%, 97/259), or mobile phone apps (37.2%, 96/259). However, a majority of participants would join an online support group (69.0%, 179/259), join a Facebook support group (62.3%, 161/259), sign up to receive text messages (71.9%, 186/259), or use an app placed on their mobile phone (70.4%, 182/259) to support their personal recovery. These differences were not significant among generations.

Discussion

Principal Findings

The results shown here demonstrate that technology adoption and internet use continues to rise, even among populations with substance use disorders. Previous research has shown that technological interventions using text message features [25] for substance using populations could be of benefit, though here we show that interventions delivered on social media platforms may be preferential. In addition, the “digital divide” that has been previously alluded to in the study of technological interventions in patients with substance use disorder [25], has been described to exist largely across racial lines. However, our results show that ethnic minorities present with rates similar to that of the general population in regards to mobile phone ownership and technology/internet use. The advent of mobile phones with capabilities has likely assisted in the partial bridging of this digital divide, also referenced as “digital differentiation” [26], supported by the prevalence of mobile phone ownership with app capabilities presented here (64.1%, 166/259).

The increased availability and use of social media platforms should also be viewed as potentially harmful to populations engaging in substance use disorder treatment. The majority of respondents in the current study had at least sometimes seen information that had resulted in the desire to return to substance use. Patients engaged in an outpatient setting that have regular access to social media and other digital platforms are at greater risks of encountering this information. The reported risk of relapse of patients in outpatient treatment settings has been as high as 70% [23]. With the risk of relapse for these populations already being high, the prevalence of those encountering triggering information on social media is high enough—47.4% (90/190) reported here—to support the increase in use of mediating supports, either in the form of recovery related information on similar platforms or targeted interventions using digital media.

Our results support that a recovery-focused social network may prove beneficial, especially to younger populations. Though this finding is age-specific, the Millennials and Generation Xers will soon make up the majority of people in substance use disorder treatment, suggesting that interventions and support services curtailed to this milieu is critical to positively improving treatment and recovery outcomes in the long-term.

Treatment centers offering substance use disorder specific or ancillary services should continue to inform themselves of potential benefits and harms of digital platforms as technology use and ownership continues to increase in all segments of the population, even those in lower socio-economic brackets. Factors such as these that have the potential to impact relapse vulnerability, outreach mechanisms, treatment engagement, and continuing aftercare should be discussed at length in service provision in the 21st century.

Limitations

The sample consisted predominantly of black, low socio-economic status males. Though the sample is not reflective of the general population, the results are comparable with other technology use and ownership studies that have shown similar prevalence rates among other demographic cross-sections—while showing lower prevalence among minority communities. The results shown here suggest that technology adoption has increased exponentially among lower socio-economic status minority communities as it has become cheaper and more readily available. A secondary limitation is the geographical location of participants coming from one metropolitan area in the northeastern United States. Results are likely not generalizable to more rural areas of the United States, and the study should be replicated with a sample representative of these areas to confirm technology adoption and impacts on treatment and recovery from substance use disorders. In addition, the current study did not clearly delineate the types of self-defined information seen by respondents in regards to emotionally triggering or recovery-related information. Future studies should seek to identify the types of channels and format this information takes so that future targeted interventions can be better informed.

Conclusions

Technology has continued to be adopted and used at increasing rates among all sectors of the population, including lower socio-economic status African-Americans. Seemingly ubiquitous mobile phone ownership and social media use among younger generations suggest that these platforms can have an immediate impact—potentially detrimental or beneficial—on an individual's treatment and recovery from a substance use disorder. Substance use disorder treatment providers should consider the implications of technology ownership and digital media use in the modification of treatment protocols, where recovery-focused platforms can be used to impact relapse vulnerability, treatment engagement, and long-term recovery outcomes. Similarly, implementing provisions to mitigate the risk of drug-related cravings resulting from seeing drug-related information on digital media platforms should also be considered, especially for younger generation clients.

Acknowledgments

The study was funded by National Institute on Drug Abuse grant R01 DA 039457.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Survey instrument.

[PDF File (Adobe PDF File), 366KB - jmir_v20i3e84_app1.pdf]

Multimedia Appendix 2

Odds ratio, generational differences in technology (*P<.05).

[PNG File, 67KB - jmir_v20i3e84_app2.png]

References

1. Mobile fact sheet. Washington, DC: Pew Research Center Internet & Technology; Feb 05, 2018.
2. Fischer HH, Moore SL, Ginosar D, Davidson AJ, Rice-Peterson CM, Durfee MJ, et al. Care by cell phone: text messaging for chronic disease management. Am J Manag Care 2012 Feb 01;18(2):e42-e47 [FREE Full text] [Medline: 22435883]
3. Fjeldsoe BS, Marshall AL, Miller YD. Behavior change interventions delivered by mobile telephone short-message service. Am J Prev Med 2009 Feb;36(2):165-173. [doi: 10.1016/j.amepre.2008.09.040] [Medline: 19135907]
4. Krishna S, Boren S, Balas E. Healthcare via cell phones: a systematic review. Telemed J E Health 2009 Apr;15(3):231-240. [doi: 10.1089/tmj.2008.0099] [Medline: 19382860]
5. Logan AG. Transforming hypertension management using mobile health technology for telemonitoring and self-care support. Can J Cardiol 2013 May;29(5):579-585. [doi: 10.1016/j.cjca.2013.02.024] [Medline: 23618506]
6. Quinn CC, Clough SS, Minor JM, Lender D, Okafor MC, Gruber-Baldini A. WellDoc mobile diabetes management randomized controlled trial: change in clinical and behavioral outcomes and patient and physician satisfaction. Diabetes Technol Ther 2008 Jun;10(3):160-168. [doi: 10.1089/dia.2008.0283] [Medline: 18473689]
7. Bonevski B, Randell M, Paul C, Chapman K, Twyman L, Bryant J, et al. Reaching the hard-to-reach: a systematic review of strategies for improving health and medical research with socially disadvantaged groups. BMC Med Res Methodol 2014 Mar 25;14:42 [FREE Full text] [doi: 10.1186/1471-2288-14-42] [Medline: 24669751]
8. Lim M, Hocking J, Hellard M, Aitken C. SMS STI: a review of the uses of mobile phone text messaging in sexual health. Int J STD AIDS 2008 May;19(5):287-290. [doi: 10.1258/ijsa.2007.007264] [Medline: 18482956]

9. Glenn T, Monteith S. New measures of mental state and behavior based on data collected from sensors, smartphones, and the Internet. Curr Psychiatry Rep 2014 Dec;16(12):523. [doi: 10.1007/s11920-014-0523-3] [Medline: 25308392]
10. Payne H, Lister C, West J, Bernhardt J. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
11. Baker TB, Gustafson DH, Shah D. How can research keep up with eHealth? Ten strategies for increasing the timeliness and usefulness of eHealth research. J Med Internet Res 2014 Feb 19;16(2):e36 [FREE Full text] [doi: 10.2196/jmir.2925] [Medline: 24554442]
12. Torous J, Powell AC. Current research and trends in the use of smartphone applications for mood disorders. Internet Interv 2015 May;2(2):169-173. [doi: 10.1016/j.invent.2015.03.002]
13. Torous J, Staples P, Onnela J. Realizing the potential of mobile mental health: new methods for new data in psychiatry. Curr Psychiatry Rep 2015 Aug;17(8):602 [FREE Full text] [doi: 10.1007/s11920-015-0602-0] [Medline: 26073363]
14. 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 Feb 07;13:49 [FREE Full text] [doi: 10.1186/1471-244X-13-49] [Medline: 23391304]
15. 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 [FREE Full text] [doi: 10.1093/schbul/sbu033] [Medline: 24609454]
16. Faurholt-Jepsen M, Frost M, Vinberg M, Christensen E, Bardram J, Kessing L. Smartphone data as objective measures of bipolar disorder symptoms. Psychiatry Res 2014 Jun 30;217(1-2):124-127. [doi: 10.1016/j.psychres.2014.03.009] [Medline: 24679993]
17. Pramana G, Parmanto B, Kendall P, Silk J. 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]
18. BinDhim N, Shaman A, Trevena L, Basyouni M, Pont L, Alhawassi T. Depression screening via a smartphone app: cross-country user characteristics and feasibility. J Am Med Inform Assoc 2015 Jan;22(1):29-34 [FREE Full text] [doi: 10.1136/amiajnl-2014-002840] [Medline: 25326599]
19. Burns MN, Begale M, Duffecy J, Gergle D, Karr CJ, Giangrande E, et al. Harnessing context sensing to develop a mobile intervention for depression. J Med Internet Res 2011 Aug 12;13(3):e55 [FREE Full text] [doi: 10.2196/jmir.1838] [Medline: 21840837]
20. Gustafson DH, McTavish F, Chih M, Atwood A, Johnson R, Boyle M, et al. A smartphone application to support recovery from alcoholism: a randomized clinical trial. JAMA Psychiatry 2014 May;71(5):566-572 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.4642] [Medline: 24671165]
21. Substance Abuse Mental Health Services Administration. HHS Publication No. SMA–17–BaroUS–16. Rockville, MD: Substance Abuse and Mental Health Services Administration; 2015. Behavioral Health Barometer: United States, Volume 4: Indicators as measured through the 2015 National Survey on Drug Use and Health and National Survey of Substance Abuse Treatment Services URL: https://store.samhsa.gov/shin/content/SMA17-BAROUS-16/SMA17-BAROUS-16.pdf [WebCite Cache ID 6xQw7Gq4o]
22. Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality. DASIS Series S-61, HHS Publication No. (SMA) 12-4701. Rockville, MD: Substance Abuse and Mental Health Services Administration; 2012. Treatment Episode Data Set (TEDS): 2000-2010. National Admissions to Substance Abuse Treatment Services URL: https://wwwdasis.samhsa.gov/dasis2/teds_pubs/2010_teds_rpt_natl.pdf [WebCite Cache ID 6xQwZBRmc]
23. Sinha R. New findings on biological factors predicting addiction relapse vulnerability. Curr Psychiatry Rep 2011 Oct;13(5):398-405 [FREE Full text] [doi: 10.1007/s11920-011-0224-0] [Medline: 21792580]
24. Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality. BHSIS Series S-91, HHS Publication No. (SMA) 17-5037. Rockville, MD: Substance Abuse and Mental Health Services Administration; 2017. Treatment Episode Data Set (TEDS): 2005-2015. National Admissions to Substance Abuse Treatment Services URL: https://wwwdasis.samhsa.gov/dasis2/teds_pubs/2015_teds_rpt_natl.pdf [accessed 2018-02-19] [WebCite Cache ID 6xQxKtv8P]
25. McClure EA, Acquavita SP, Harding E, Stitzer ML. Utilization of communication technology by patients enrolled in substance abuse treatment. Drug Alcohol Depend 2013 Apr 01;129(1-2):145-150 [FREE Full text] [doi: 10.1016/j.drugalcdep.2012.10.003] [Medline: 23107600]
26. Riddlesden D, Singleton AD. Broadband speed equity: a new digital divide? Appl Geogr 2014 Aug;52:25-33. [doi: 10.1016/j.apgeog.2014.04.008]

Abbreviations

AA: Alcoholics Anonymous
NA: Narcotics Anonymous

Edited by G Eysenbach; submitted 13.10.17; peer-reviewed by A Brown, N Bragazzi, N Unciano; comments to author 23.11.17; revised version received 26.11.17; accepted 14.12.17; published 06.03.18.

Please cite as:
Ashford RD, Lynch K, Curtis B
Technology and Social Media Use Among Patients Enrolled in Outpatient Addiction Treatment Programs: Cross-Sectional Survey Study
J Med Internet Res 2018;20(3):e84
URL: http://www.jmir.org/2018/3/e84/
doi:10.2196/jmir.9172
PMID:29510968

Original Paper

Growth Dynamics of Patient-Provider Internet Communication: Trend Analysis Using the Health Information National Trends Survey (2003 to 2013)

Will L Tarver[1], MLIS, DrPH; Terri Menser[2], MBA, PhD; Bradford W Hesse[3], PhD; Tyler J Johnson[4], MPH, MSBS; Ellen Beckjord[5], MPH, PhD; Eric W Ford[6], MPH, PhD; Timothy R Huerta[4], MS, PhD

[1]Health Services Research and Development Service Center for Health Information and Communication, Richard L Roudebush VA Medical Center, Indianapolis, IN, United States

[2]Center for Outcomes Research, Houston Methodist Research Institute, Houston, TX, United States

[3]Health Communication and Informatics Research Branch, National Cancer Institute, Bethesda, MD, United States

[4]Department of Family Medicine, The Ohio State University, Columbus, OH, United States

[5]Population Health Program Design and Engagement Optimization, UPMC Health Plan, Pittsburgh, PA, United States

[6]Department of Health Care Organization and Policy, University of Alabama at Birmingham, Birmingham, AL, United States

Corresponding Author:
Timothy R Huerta, MS, PhD
Department of Family Medicine
The Ohio State University
2231 N High Street
Columbus, OH, 43201
United States
Phone: 1 614 293 8973
Fax: 1 614 293 2715
Email: tim.huerta@osumc.edu

Abstract

Background: Communication is key in chronic disease management, and the internet has altered the manner in which patients and providers can exchange information. Adoption of secure messaging differs among patients due to the digital divide that keeps some populations from having effective access to online resources.

Objective: This study aimed to examine the current state of online patient-provider communication, exploring trends over time in the use of online patient-provider communication tools.

Methods: A 3-part analytic process was used to study the following: (1) reanalysis, (2) close replication across years, and (3) trend analysis extension. During the reanalysis stage, the publicly available Health Information National Trends Survey (HINTS) 1 and 2 data were used with the goal of identifying the precise analytic methodology used in a prior study, published in 2007. The original analysis was extended to add 3 additional data years (ie, 2008, 2011, and 2013) using the original analytical approach with the purpose of identifying trends over time. Multivariate logistic regression was used to analyze pooled data across all years, with year as an added predictor, in addition to a model for each individual data year.

Results: The odds of internet users to communicate online with health care providers was significantly and increasingly higher year-over-year, starting in 2003 (2005: odds ratio [OR] 1.31, 95% CI 1.03-1.68; 2008: OR 2.14, 95% CI 1.76-2.59; 2011: OR 2.92, 95% CI 2.33-3.66; and 2013: OR 5.77; 95% CI 4.62-7.20). Statistically significant socio-economic factors found to be associated with internet users communicating online with providers included age, having health insurance, having a history of cancer, and living in an urban area of residence.

Conclusions: The proportion of internet users communicating online with their health care providers has significantly increased since 2003. Although these trends are encouraging, access challenges still exist for some groups, potentially giving rise to a new set of health disparities related to communication.

(J Med Internet Res 2018;20(3):e109) doi:10.2196/jmir.7851

KEYWORDS

internet; health communication

Introduction

Effective communication between a patient and their health care provider is central to the provision of medical care [1-3]. Without effective communication, and the trust it can build, chronic disease management can be tenuous and the delivery of high quality care is impaired [4], medication adherence is reduced, and patients rely on low-quality information for making decisions that affect their health [5]. The internet has transformed the way people communicate, providing alternative means of communication (eg, email, secure messaging, instant messaging, and online videos) that supplement or, in some instances, have replaced the traditional in-person and telephonic communications. These alternative communication channels are being utilized with greater frequency and are becoming a normal part of service delivery in the health care industry. Empirical studies have found a positive correlation between increased electronic communication and health outcomes.

The use of internet-enabled communication can facilitate patient engagement and create better documentation modes for patient-provider communication [6,7]. There is a growing interest in shifting care processes (eg, requests for referral, test results) to technology-enabled models. For physicians, decreasing face-to-face consultation time for low-value administrative activities allows them to focus on more important clinical encounters [6,7]. For patients, the new communication channels can reduce access burdens related to navigating the care system, including the transaction costs associated with seeking care (eg, transportation, taking time off of work to seek health care). A recent study by Reed and colleagues [8] found that patients with higher out-of-pocket costs were significantly more likely to use secure email as their first method of contact with respect to their health care. In addition, many patients who used a secure email to communicate with their health care providers reported that it reduced their phone contacts and/or office visits.

The use of electronic communication with patients provides important opportunities for the provision of patient-centered care [9]. Electronic communication has been found to improve patient satisfaction and saves patient time. However, providers who use electronic communication extensively with their patients have noted there is a tradeoff that needs to occur if efficiency gains are to be realized. In particular, extensive email communication increases doctors' workload unless the number of office visits per patient is reduced [10]. A systematic review of patient-provider email communication found the benefits of electronic communication to be recognized by patients and providers alike (eg, ease of communication and ability to improve health care), but several included studies also identified barriers to its use including workload and time demands, confidentiality and security, lack of reimbursement, and inappropriate use of email by patients [11]. One policy change these studies suggest is that insurers should reimburse for electronic communication to promote online patient engagement. Electronic communication is an avenue that may be able to effectively address basic patient questions, leaving the patient visit to focus on more critical issues and concerns.

Despite early literature showing that patients were receptive to communicating through email with their providers [12], significant challenges to both patients and providers have prevented its widespread use (eg, the digital divide, referring to the lack of equity in availability of technologies by demographic characteristics and geographic location [6]; lack of physician reimbursement [6]; and mixed evidence in support of electronic communication on health outcomes [13,14]). In 2007, Beckjord and colleagues [15] used the Health Information National Trends Survey (HINTS) in a benchmark study reporting the prevalence of, and factors related to, use of online communication between patients and their providers. The study found a low prevalence of online patient-provider communication, which significantly increased, from 7% of internet users to 10% between the years 2003 and 2005. In addition, factors that influenced online communication among internet users included higher education, living in a metro area, having poorer health status, and having a personal history of cancer. Since this publication, the growth of the internet has changed the landscape of how people search for and receive messages regarding health and their health status (health information seeking) [16] and how they communicate electronically within an online format (online communication) [17] in the health arena. For example, the internet is used to gather health information by 66% of adults with no chronic conditions and 51% of those with chronic conditions [18]. Although overall access to the Web has increased greatly, with 84% of American adults using the internet in 2015, an increase of nearly 65% from 2000 when 51% of the population used the internet [19]; the digital divide continues to be an issue. Home internet availability is lower among older populations, racial or ethnic minorities, less educated groups, those with lower incomes, and for people that reside outside of metropolitan areas [20-22]. However, increased internet access has also been seen in the underserved populations through the proliferation of mobile devices, which make the internet more accessible [22]. The widespread use of mobile devices has also influenced the way people communicate. For example, electronic communications such as text messaging and email are being used by more than 90% of the population [23].

The Diffusion of Innovation Theory discusses the process by which a new product or technology is adopted by a given population [24]. Understanding the diffusion of online communication can inform the ways in which providers interact with patients. Over half of US health care providers (57%) reported having a patient portal in place in 2012 [25]. The adoption of patient portals has been encouraged by federal policy initiatives such as Meaningful Use; a core objective for Stage 2 Meaningful Use is to use secure electronic messaging to communicate with patients on relevant health information to impact patient care and safety [26]. Kannry and colleagues noted that there is an opportunity to increase secure messaging, citing Ralston and colleagues' study, which indicated that nearly a

third of outpatient encounters could be conducted via secure messaging [27,28]. New, anecdotal evidence suggests that the use of patient portals for communication can help lower overhead costs by reducing the patient call volume and decreasing the overall time required to communicate with patients [29]. Despite these technological advancements, little is known about the current state of online patient-provider communication and how its use has changed over time. Such data will be valuable in informing us about the future direction of online patient-provider communication.

The objectives of this paper were to (1) reproduce and replicate the initial study conducted by Beckjord and colleagues [15], (2) examine the current state of online patient-provider communication, and (3) explore trends over time in the use of online patient-provider communication. We expand on the original study by utilizing data from 5 iterations of the HINTS dataset to explore whether online patient-provider communication and the sociodemographic factors influencing its use have changed over time. Samples were weighted to make the samples representative of the US population.

Methods

Data

Although HINTS is a dataset that is maintained by National Cancer Institute and does include several questions that focus on cancer, the population surveyed is random and many of the questions relate to general health communication. HINTS has 6 iterations, 5 of which include the requisite variables to extend Beckjord and colleagues' study. The first iteration of HINTS (HINTS 1) was conducted from October 2002 through April 2003. Data collection was achieved via a random digit dial telephone survey, which generates phone numbers at random. Selected phone numbers were submitted to a telephone matching service in an effort to remove nonresidential numbers. A computer-assisted telephone interview (CATI) format was utilized to accommodate complex skip-patterns. Survey administration averaged 30 min per respondent, and data were collected from 6369 respondents.

HINTS 2 (2005) data were collected from February 2005 through August 2005. Like HINTS 1, HINTS 2 also used a random digit dialing telephone survey with a CATI format. All data collection procedures were identical to HINTS 1. Data were collected from 5586 respondents.

The HINTS 3 (2008) data were collected from January through May 2008. The HINTS 3 sample design included 2 data collection methods. One sample was drawn as a random digit dialing telephone survey, using a CATI format. The second national random sample was selected from a list of addresses from the United States Postal Service (USPS) administrative records. Data were collected from 4092 respondents via CATI and from 3582 respondents via mail. Differences between the CATI and USPS samples were tested. Despite the differences between the 2 populations in education, income, general health status, metropolitan statistical area, and internet use, all data were included to maintain fidelity to the original study.

HINTS 4, Cycles 1 and 3 data were collected from October 2011 through February 2012 and September through December 2013, respectively. The sample design for both surveys consisted of a single-mode mail survey, using the Next Birthday Method for respondent selection. The sample design for the interview portion of the survey consisted of 2 stages. In the first stage, a stratified sample of addresses was selected from a file of residential addresses. In the second stage, 1 adult was selected within each sampled household. The sampling frame consisted of a database of addresses used by Marketing Systems Group to provide random samples of addresses. Complete data were collected from 3959 respondents in 2011 and from 3185 respondents in 2013.

Measures

Sociodemographic characteristics included age (18-34, 35-49, 50-65, or >65 years), sex (male or female), education (less than high school, high school graduate, some college, or college graduate), income (<US $10,000; US $10,000-14,999; US $15,000-19,999; US $20,000-34,999; US $35,000-49,999; US $50,000-74,999; or ≥US $75,000), race and ethnicity (white, Hispanic, black, or Asian, other), metropolitan area (metro or non-metro), and health insurance status (yes or no), one's own perceived health status (excellent, very good, good, fair, or poor), and cancer history (yes or no). Measures relating to use of technology and electronic communication were assessed in the form of questions, which asked whether they used the internet (yes or no) and whether respondents had communicated online with a provider in the past 12 months (yes or no).

There were differences in how data were categorized for the income and perceived health status variables. The categories for the income variable in HINTS changed over time, and the income categories used in this study of trends over time reflect that. Additionally, the categories for perceived health status remained stratified for a more in-depth analysis, and were not dichotomized as in the original study.

Analytic Framework

A 3-part analytic process was used to study (1) reanalysis, (2) close replication across years, and (3) trend analysis extension (see Multimedia Appendix 1). During the reanalysis stage, the publicly available HINTS 1 and 2 data were used with the goal of identifying the precise analytic methodology used in the original 2007 paper. This stage served as a precursor to replication, ensuring that the model with additional data years is done with a high degree of fidelity to the analytic framework established as valid in the literature [30]. In the close replication stage, we made minor alterations to the original analytical model to enable effective analysis of multiple data years in a manner that is as close to the original model as possible, while taking into account all available variables in each data year. Finally, we engaged in extension, that is, the analysis of additional data years, 2008, 2011, and 2013, using the same analytical approach with the purpose of identifying trends over time. The first author of the original study was invited to join the study at this stage.

Statistical Analyses

Analyses were conducted using both unweighted and weighted data. All weighted analyses were conducted using jackknife

variance estimation to account for the complex survey design and provide nationally representative population estimates. Mirroring the original paper [15], respondents' sociodemographic characteristics were calculated by percentage of internet users for the years 2003, 2005, 2008, 2011, and 2013 (Table 1).

The same characteristics were then calculated by the percentage of patients who communicated online with a provider (Table 2). We employed a weighted, multivariate logistic regression for each individual data year (Multimedia Appendix 2), as well as an additional model, which pooled the data across years (Table 3) and contained the year as a predictor variable to identify potential predictors of online communication with a health care provider. Each model was adjusted to control for all sociodemographic characteristics and assessed online communication with a provider where responses of "Yes" represented a positive outcome. The original study contained a table of the results of the bivariate year-specific analyses, but trying to assess changing trends over time related to demographic factors may present spurious correlations as associations; thus, we omitted this approach in this study.

Results

Sample Characteristics and Bivariate Analyses

The sample characteristics of internet users for the reproduced Beckjord and colleagues [15] study using HINTS 2003 and 2005 data, and extending it to include the 2008, 2011, and 2013 HINTS iterations are presented in Table 1.

In 2003, 7% of internet users reported communicating online with a health care provider in the past 12 months, increasing to nearly 10% of internet users in 2005. These estimates are consistent with the original study, and the percentage of internet users reported communicating online with health care providers increased to 14% in 2008, 19% in 2011, and 30% in 2013.

Consistent with Beckjord and colleagues' original study, individuals who communicated online with health care providers had significantly more years of education and were more likely to reside in a metropolitan area in 2003. In 2005, these individuals were more likely to be female and were more likely to have a personal history of cancer. The reproduction of the original study found minor differences in the means of some variables, which may be attributed to variations in the SUDAAN (Research Triangle Park, North Carolina) and STATA (College Station, Texas) jackknife algorithms. In addition, we found 2 differences between sociodemographic variables and communicating online with health care providers for 2005 when comparing results with those from the original study (see Table 2). These 2 differences were (1) communicating online with health care providers was not found to be statistically significantly associated with more years of education (P>.05) and (2) annual income was statistically significant (P=.038).

Multivariate Analyses

The findings of our multivariate models conflicted with the original study [15] in 2 instances in 2005. First, gender was not found to be statistically significant, and second, we found that individuals who reported poor health had higher odds of communicating online with health care providers, which was not reported in the original study. Notably, several variables were consistently found significant across multiple years. For example, gender was statistically significant for the years 2011 and 2013; having a history of cancer was statistically significant for 2005 and 2013; and geographic location was statistically significant for 2003, 2008, and 2011 (all P values were <.05). Full results of multivariate analyses by HINTS year is provided in Multimedia Appendix 2.

When exploring trends in the use of online communication with health care providers, we found a significant increase between the years 2003 and 2013. More specifically, as compared with 2003, the odds of internet users to communicate online with health care providers was found to be significantly and increasingly higher in the subsequent years, 2005 (odds ratio [OR] 1.31, 95% CI 1.03-1.68, P=.027), 2008 (OR 2.14, 95% CI 1.76-2.59, P<.001), 2011 (OR 2.92, 95% CI 2.33-3.66, P<.001), and 2013 (OR 5.77, 95% CI 4.62-7.21, P<.001).

When looking at all 5 years of HINTS data and adding in year as a covariate, the odds of communicating with a provider via the internet increased for women (OR 1.31, 95% CI 1.10-1.55, P=.002), college graduates (OR 1.88, 95% CI 1.08-3.28, P=.026), and data year (see Table 3). Conversely, the odds of communicating with a provider via the internet decreased for (1) individuals between 65 and 74 years of age (OR 0.70, 95% CI 0.52-0.94, P=.018), (2) the uninsured (OR 0.59, 95% CI 0.41-0.85, P=.005), (3) individuals with no history of cancer (OR 0.68, 95% CI 0.54-0.84, P=.001), and (4) individuals living in a nonmetro area (OR 0.63, 95% CI 0.49-0.80, P<.001).

Table 1. Weighted percentage of Health Information National Trends Survey (HINTS) internet users.

Characteristic	HINTS 2003 (n=3982), %	HINTS 2005 (n=3244), %	HINTS 2008 (n=5078), %	HINTS 2011 (n=2914), %	HINTS 2013 (n=2284), %
Communicated online with a provider in the past 12 months	7.00	9.63	13.57	19.11	29.70
Age, years					
18-34	38.29	37.74	36.16	34.89	31.36
35-49	35.77	33.25	32.81	30.33	32.63
50-64	19.82	22.09	23.03	24.67	25.23
65-74	4.37	4.97	5.49	6.99	7.13
75 or older	1.76	1.95	2.51	3.12	3.64
Gender					
Male	49.50	48.22	47.13	47.81	48.87
Education					
Less than high school	6.84	5.30	5.39	7.25	5.14
High school graduate	25.51	23.58	21.76	18.86	19.60
Some college	32.98	38.28	40.37	34.45	36.53
College graduate	34.67	32.84	32.48	39.45	38.73
Annual income, US $					
<10,000	2.76	3.20	4.35	7.78	4.78
10,000 to <15,000	2.14	3.26	4.50	5.17	4.06
15,000 to <20,000	4.45	3.12	3.18	6.53	5.27
20,000 to <35,000	17.45	13.20	12.59	15.51	12.37
35,000 to <50,000	18.32	13.88	14.38	12.30	15.29
50,000 to <75,000	22.34	25.52	21.93	18.42	19.71
75,000 or more	32.53	37.82	39.06	34.29	38.52
Race/ethnicity					
White	78.16	76.57	74.74	68.71	70.36
Hispanic/Latino	6.91	7.55	9.06	13.68	13.03
African American	8.71	8.90	9.35	10.34	9.36
Asian	3.10	3.24	4.96	5.34	4.85
Other	3.13	3.74	1.90	1.94	2.40
Health insurance					
Yes	89.32	87.50	86.01	82.83	84.56
Health status					
Excellent	15.31	13.35	11.84	15.19	13.98
Very good	34.24	33.27	40.42	39.56	38.98
Good	34.62	36.98	36.29	32.80	36.71
Fair	13.12	13.78	9.70	10.48	9.26
Poor	2.72	2.62	1.75	1.97	1.07
History of cancer					
Yes	8.42	9.12	5.96	6.98	7.46
Metropolitan statistical area					
Metro area county	84.38	81.79	85.13	84.88	83.24

Table 2. Weighted percentage of patients who communicated online with a health care provider. Statistically significant values are italicized. HINTS: Health Information National Trends Survey.

Characteristic	HINTS 2003, %	*P* value	HINTS 2005, %	*P* value	HINTS 2008, %	*P* value	HINTS 2011, %	*P* value	HINTS 2013, %	*P* value
Age in years		.11		.66		.19		.26		.16
18-34	6.36		10.25		12.42		18.13		27.7	
35-49	6.75		9.27		15.64		17.34		33.13	
50-64	9.32		10.09		12.94		20.04		30.41	
65-74	4.62		6.45		14.14		20.82		22.18	
75 or older	6.72		7.01		11.05		29.38		21.24	
Gender		.17		*.03* [a]		.98		*.02* [a]		.21
Male	7.64		7.90		13.62		15.97		27.58	
Female	6.38		11.24		13.58		22.12		31.87	
Education		*<.001* [b]		.145		*<.001* [b]		*<.001* [b]		*<.001* [b]
Less than high school	3.17		8.53		10.80		8.18		17.20	
High school graduate	3.53		6.60		9.74		10.37		15.25	
Some college	7.28		10.09		11.44		17.52		29.06	
College graduate	10.31		11.66		19.66		25.74		39.42	
Annual income (US $)		.32		*.04* [a]		*.007* [b]		*.04* [a]		*<.001* [b]
<$10,000	8.51		16.01		11.68		15.12		17.91	
$10,000 to <$15,000	9.32		7.02		13.50		9.43		23.86	
$15,000 to <$20,000	8.97		3.84		10.64		14.04		14.46	
$20,000 to <$35,000	6.43		7.85		8.71		12.36		13.90	
$35,000 to <$50,000	5.53		7.57		11.54		17.62		26.58	
$50,000 to <$75,000	6.51		8.43		12.05		17.75		30.59	
$75,000 or more	9.16		12.78		19.72		25.36		41.14	
Race/ethnicity		.96		.36		.99		.16		.18
White	7.28		9.52		13.80		19.75		29.55	
Hispanic/Latino	6.42		5.93		14.15		17.05		29.05	
African American	6.14		11.33		13.02		12.35		31.74	
Asian	7.08		11.09		13.94		26.07		45.47	
Other	7.36		15.59		15.18		30.78		25.21	
Health insurance		.56		.30		*.003* [b]		*.001* [b]		*.003* [b]
Yes	7.25		10.08		14.68		21.27		32.58	
No	6.26		7.83		6.74		8.54		15.65	
Health status		.59		.20		.27		.59		.16
Excellent	7.57		7.89		16.60		22.39		38.12	
Very good	6.55		9.83		13.57		17.25		26.59	
Good	6.83		10.61		12.65		20.73		30.60	
Fair	8.66		7.81		15.60		15.80		24.49	
Poor	6.49		17.85		8.58		22.34		31.23	
History of cancer		.11		*.01* [a]		.16		.12		*.04* [a]
Yes	9.37		14.60		16.44		23.82		39.88	
No	6.80		9.14		13.53		18.75		29.01	

Characteristic	HINTS 2003, %	*P* value	HINTS 2005, %	*P* value	HINTS 2008, %	*P* value	HINTS 2011, %	*P* value	HINTS 2013, %	*P* value
Metropolitan statistical area		*.002* [b]		.14		*.002* [b]		*.001* [b]		*.002* [b]
Metro area county	7.51		10.11		14.20		20.65		31.72	
Nonmetro area county	4.28		7.46		9.90		10.42		19.56	

[a]*P*<.05.
[b]*P*<.01.

Table 3. Odds of communicating online with a health care provider within past 12 months across years (n=14,446). Statistically significant values are italicized. HINTS: Health Information National Trends Survey.

Characteristic	HINTS 2003-2013, odds ratio (95% CI)	*P* value
Age in years	Reference: 18-34	
35-49	0.97 (0.78-1.20)	.80
50-64	0.93 (0.75-1.17)	.54
65-74	*0.70 (0.52-0.94)*	*.02* [a]
75 or older	1.16 (0.78-1.73)	.47
Gender	Reference: male	
Female	*1.31 (1.10-1.55)*	*.002* [b]
Education	Reference: less than high school	
High school graduate	0.83 (0.48-1.46)	.52
Some college	1.43 (0.82-2.49)	.20
College graduate	*1.88 (1.08-3.28)*	*.03* [a]
Annual income (US $)	Reference: <US $10,000	
$10,000 to <$15,000	1.17 (0.55-2.49)	.68
$15,000 to <$20,000	0.84 (0.38-1.86)	.67
$20,000 to <$35,000	0.87 (0.44-1.71)	.69
$35,000 to <$50,000	1.17 (0.61-2.25)	.64
$50,000 to <$75,000	1.27 (0.68-2.38)	.45
$75,000 or more	1.75 (0.94-3.23)	.08
Race/ethnicity	Reference: white	
Hispanic/ Latino	1.04 (0.77-1.40)	.81
African American	1.00 (0.75-1.32)	.98
Asian	1.22 (0.85-1.76)	.29
Other	1.21 (0.82-1.78)	.34
Health insurance	Reference: yes	
No	*0.59 (0.41-0.85)*	*.005* [b]
History of cancer	Reference: Yes	
No	*0.68 (0.54-0.84)*	*.001* [b]
Metropolitan statistical area	Reference: metro area county	
Nonmetro area county	*0.63 (0.49-0.80)*	*<.001* [b]
Data year	Reference: 2003	
2005	*1.31 (1.03-1.68)*	*.03* [a]
2008	*2.14 (1.76-2.59)*	*<.001* [b]
2011	*2.92 (2.33-3.66)*	*<.001* [b]
2013	*5.77 (4.62-7.20)*	*<.001* [b]

[a]*P*<.05.

[b]*P*<.01.

Discussion

Principal Findings

Beckjord and colleagues [15] reported the prevalence of and changes in the use of online patient-provider communication in 2003 and 2005 and described sociodemographic and health-related factors associated with its use. In this study, we were able to use subsequent iterations of the HINTS dataset to describe the current state of online patient-provider communication and understand how its use has changed over time. The data year is shown to be statistically significant in all years with the odds of communicating with a provider online more than quadrupling from 2003 to 2013, the latter indicative of the diffusion of technology over time. This finding coincided with a recent study showing a significant increase in surgeons' use of secure messaging; the proportion of outpatient interactions was 5.4% in 2008 compared with 15.3% in 2010 [31].

Although we were able to match the majority of the coefficients in the re-analysis stage, there were some minor differences, which may have been attributable to differences in the jackknife algorithm deployed in the software packages used. That said, the interpretation and findings of the first year (2003) did not change as a result; however, several changes were experienced in the 2005 analysis. These differences may be partially attributed to higher estimates of internet use among low-income individuals in 2005. These estimates are not surprising, given the literature reporting interest among underserved populations in using electronic and internet-based tools to communicate with their providers [32]. In addition, evidence shows increasing adoption of patient portals among providers serving underserved populations in efforts to increase patient engagement [33]. The health disparity literature brings attention to the importance of the quality of patient-provider communication when providing medical care to minority populations [34] and its contribution to disparities in medical care [35,36].

We also found that the use of online patient-provider communication steadily increased over time between the years 2003 and 2013. This upsurge in use may be largely attributed to increased interest in the use of online communication tools [32]. In addition, recent attention has been given to using online communication with providers as a means to provide patient-centered care [37] while federal policy initiatives require that physicians provide the use of secure electronic messaging to their patients through Meaningful Use Stage 2 requirements [27]. This federal push to adopt electronic health records has also led to an increase in the implementation and availability of tools to facilitate communication between patients and their providers, such as personal health records (PHRs). It is projected that 75% of adults will use PHRs by the year 2020 [38].

Beckjord and colleagues [15] also explored sociodemographic and health-related factors associated with the use of online communication with a health care provider. We found that in 2003 and 2005, women were marginally more likely to communicate online with their providers, and significantly more likely to communicate with their providers in 2011 and 2013. This is consistent with findings from a previous study identifying women as being more likely to use email to communicate with their doctors [39]. The trends may be explained by understanding gender differences in use of the internet. According to the Pew Research Center [19], men were more likely than women to use the internet in the early 2000s. However, this gap gradually decreased over the years and became equal around the year 2008. In addition, there is evidence that women are more likely than men to use the internet to access health information and to get support for health problems [40]. A recent study looking at patient portal use found that education and sex remain statistically significant when controlling for internet access and preference of communication mode (in person or over the phone vs patient portal) [41].

Being located in a metropolitan area is also shown as relatively consistent when it comes to communicating online with a health care provider. This may be partially attributed to the digital divide. There is literature that suggests patients located in urban locations are more likely to use electronic messaging to communicate with their providers [39], whereas physicians located in urban areas are more likely to provide secure messaging to their patients [42].Finally, having a history of cancer was also found to be associated with online patient-provider communication. Due to the complexity of cancer care, patients often leave visits with their physicians overwhelmed with information and confused about their condition and treatment [43]. Literature suggests that patients need multi-level communication to facilitate information exchange and foster patient-provider relationships [44]. As a result, cancer patients benefit from an asynchronous platform that allows them to access information and engage with their health care team at their convenience. Although these data do capture general health status and cancer history, they do not include other chronic conditions. We were therefore unable to explore the relationship between chronic illnesses outside of cancer and the use of online communication with a health care provider. However, the literature does suggest that these communication technologies will also help patients to more effectively manage other chronic illnesses such as diabetes and hypertension [45]. The nonsignificant finding for health status, measured by a 5-point Likert scale, may too broadly capture respondents' health state.

Limitations

HINTS is subject to the same issues all surveys and self-report instruments are, that is, low response rates, potential sampling bias, social desirability issues particularly around issues of smoking and other lifestyle choices, and item limitations. The HINTS response rate was 33% in 2003, 21% in 2005, 21% for telephone survey and 31% for mail surveys in 2007, 37% in 2011, and 35% in 2015; response rates are not a deterministic indicator of bias [46]. We note, in looking at the multivariate models by year, that the significance of gender fluctuates depending on the data year. This could have been the result of sampling dynamics and not indicative of changes in the population at large. A study asking similar questions in the same data year would have allowed for triangulation, but no such study was available. Finally, the main outcome variable is limited by its dichotomous nature and does not allow us to identify the specific type of provider respondents communicated

online with within the previous 12 months. For example, a patient could have communicated electronically with a nurse or other member of the care team.

Implications for Future Research

As the diffusion of this technology continues, future research related to online means of patient-provider communication should remain a focus. Future studies should try and garner a better sense of the frequency of communication and document the type of provider with whom the patient communicates to inform providers and health care organizations implementing new and additional means of online communication. Additionally, understanding use by disease type, including chronic illnesses, would be a useful addition to the literature. Although the adoption of these communication technologies is increasing among minority populations, another important area to explore is the quality of communication between minorities and their providers, and how it may influence, or be influenced by, the use of these technologies. Finally, with regards to the methodological approach utilized in this study, it is important to note that previously published studies are often not updated as new waves of data become available. HINTS data have been used in nearly 400 studies, and have yielded intriguing findings, but data updates to prior studies that use large datasets are infrequent. Perhaps this is due to instrument evolution across survey cycles that makes combining data across cycles challenging, but regardless, replication and extension of prior work is an area of research that warrants further attention.

Conclusions

Despite initial challenges in implementation [6], the proportion of internet users communicating online with their health care providers has significantly increased since 2003. In addition, these trends are likely to continue with the enactment of the Health Information Technology for Economic and Clinical Health Act and Meaningful Use Stage 2 core objectives, which require providers to use secure electronic messaging to communicate with their patients. However, challenges still remain pertaining to the digital divide affecting individuals residing in nonmetropolitan areas and their access to the internet, making this group less likely to communicate online with their providers. Future research should continue to investigate patient-provider communication trends, specifically to gain an understanding of successful interventions that mitigate identified barriers from both provider and patient perspectives.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Analytical model and framework.

[PDF File (Adobe PDF File), 243KB - jmir_v20i3e109_app1.pdf]

Multimedia Appendix 2

Odds of communicating online with a health care provider in the past 12 months by year.

[PDF File (Adobe PDF File), 114KB - jmir_v20i3e109_app2.pdf]

References

1. Clark PG. Communication between provider and patient: values, biography, and empowerment in clinical practice. Ageing Soc 1996;16:747-774 [FREE Full text]
2. Ong LM, de Haes JC, Hoos AM, Lammes FB. Doctor-patient communication: a review of the literature. Soc Sci Med 1995 Apr;40(7):903-918. [Medline: 7792630]
3. Roter D, Hall J. Doctor-patient communication: Why and how communication contributes to the quality of medical care. In: Gellman MD, Turner JR, editors. Encyclopedia of Behavioral Medicine. New York, NY: Springer New York; 2013:622-627.
4. Ha JF, Longnecker N. Doctor-patient communication: a review. Ochsner J 2010;10(1):38-43 [FREE Full text] [Medline: 21603354]
5. Street Jr RL, Makoul G, Arora NK, Epstein RM. How does communication heal? Pathways linking clinician-patient communication to health outcomes. Patient Educ Couns 2009 Mar;74(3):295-301. [doi: 10.1016/j.pec.2008.11.015] [Medline: 19150199]
6. Katz SJ, Moyer CA. The emerging role of online communication between patients and their providers. J Gen Intern Med 2004 Sep;19(9):978-983 [FREE Full text] [doi: 10.1111/j.1525-1497.2004.30432.x] [Medline: 15333064]
7. Weiner JP. Doctor-patient communication in the e-health era. Isr J Health Policy Res 2012 Aug 28;1(1):33 [FREE Full text] [doi: 10.1186/2045-4015-1-33] [Medline: 22929000]
8. Reed M, Graetz I, Gordon N, Fung V. Patient-initiated e-mails to providers: associations with out-of-pocket visit costs, and impact on care-seeking and health. Am J Manag Care 2015 Dec 1;21(12):e632-e639 [FREE Full text] [Medline: 26760425]

9. Institute of Medicine. Crossing the quality chasm: A new health system for the 21st century. Washington, DC: The National Academies Press; 2001.
10. Bishop TF, Press MJ, Mendelsohn JL, Casalino LP. Electronic communication improves access, but barriers to its widespread adoption remain. Health Aff (Millwood) 2013 Aug;32(8):1361-1367 [FREE Full text] [doi: 10.1377/hlthaff.2012.1151] [Medline: 23918479]
11. Ye J, Rust G, Fry-Johnson Y, Strothers H. E-mail in patient-provider communication: a systematic review. Patient Educ Couns 2010 Aug;80(2):266-273 [FREE Full text] [doi: 10.1016/j.pec.2009.09.038] [Medline: 19914022]
12. Car J, Sheikh A. Email consultations in health care: 2--acceptability and safe application. BMJ 2004 Aug 21;329(7463):439-442 [FREE Full text] [doi: 10.1136/bmj.329.7463.439] [Medline: 15321903]
13. de Jong CC, Ros WJ, Schrijvers G. The effects on health behavior and health outcomes of Internet-based asynchronous communication between health providers and patients with a chronic condition: a systematic review. J Med Internet Res 2014;16(1):e19 [FREE Full text] [doi: 10.2196/jmir.3000] [Medline: 24434570]
14. Goldzweig CL, Orshansky G, Paige NM, Towfigh AA, Haggstrom DA, Miake-Lye I, et al. Electronic patient portals: evidence on health outcomes, satisfaction, efficiency, and attitudes: a systematic review. Ann Intern Med 2013 Nov 19;159(10):677-687. [doi: 10.7326/0003-4819-159-10-201311190-00006] [Medline: 24247673]
15. Beckjord EB, Finney Rutten LJ, Squiers L, Arora NK, Volckmann L, Moser RP, et al. Use of the internet to communicate with health care providers in the United States: estimates from the 2003 and 2005 Health Information National Trends Surveys (HINTS). J Med Internet Res 2007 Jul;9(3):e20 [FREE Full text] [doi: 10.2196/jmir.9.3.e20] [Medline: 17627929]
16. Tardy RW, Hale CL. Getting “plugged in”: a network analysis of health‐information seeking among “stay‐at‐home moms”. Commun Monogr 1998 Dec;65(4):336-357. [doi: 10.1080/03637759809376457]
17. Tu CH, Yen CJ. A study of multidimensional online social presence. In: Cooke LW, editor. Frontiers in higher education. Hauppauge, NY: Nova Publishers; 2006.
18. Pew Internet. 2010. Chronic Disease and the Internet URL: http://www.pewinternet.org/2010/03/24/chronic-disease-and-the-internet/ [accessed 2017-08-25] [WebCite Cache ID 6tvETfGR6]
19. Pew Internet. 2015. Americans’ Internet Access: 2000-2015 URL: http://www.pewinternet.org/2015/06/26/americans-internet-access-2000-2015/ [accessed 2017-08-25] [WebCite Cache ID 6pcfQNAHY]
20. Campbell RJ. Meeting seniors' information needs: using computer technology. Home Health Care Manag Pract 2008 Jan 15;20(4):328-335. [doi: 10.1177/1084822307310765]
21. Chou WY, 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]
22. Pew Internet. 2013. Who’s Not Online and Why URL: http://www.pewinternet.org/2013/09/25/whos-not-online-and-why/ [accessed 2017-08-25] [WebCite Cache ID 6pcfeo8CD]
23. Pew Internet. 2015. U.S. Smartphone Use in 2015 URL: http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/ [accessed 2017-04-25] [WebCite Cache ID 6tvDzwT8E]
24. Rogers EM. Diffusion of Innovations, 4th edition. New York, United States: Simon & Schuster; 2010.
25. Tavares J, Oliveira T. Electronic health record patient portal adoption by health care consumers: an acceptance model and survey. J Med Internet Res 2016 Mar 2;18(3):e49 [FREE Full text] [doi: 10.2196/jmir.5069] [Medline: 26935646]
26. CMS. 2012. Stage 2 Overview Tipsheet URL: https://www.cms.gov/regulations-and-guidance/legislation/ehrincentiveprograms/downloads/stage2overview_tipsheet.pdf [accessed 2018-02-19] [WebCite Cache ID 6xMXnsaw9]
27. Kannry J, Beuria P, Wang E, Nissim J. Personal health records: meaningful use, but for whom? Mt Sinai J Med 2012;79(5):593-602. [doi: 10.1002/msj.21334] [Medline: 22976365]
28. Ralston JD, Coleman K, Reid RJ, Handley MR, Larson EB. Patient experience should be part of meaningful-use criteria. Health Aff (Millwood) 2010 Apr;29(4):607-613 [FREE Full text] [doi: 10.1377/hlthaff.2010.0113] [Medline: 20368589]
29. McNeill SM. Lower your overhead with a patient portal. Fam Pract Manag 2016;23(2):21-25. [Medline: 26977985]
30. Tsang EWK, Kwan K. Replication and theory development in organizational science: a critical realist perspective. Acad Manage Rev 1999 Oct;24(4):759-780. [doi: 10.2307/259353]
31. Shenson JA, Cronin RM, Davis SE, Chen Q, Jackson GP. Rapid growth in surgeons' use of secure messaging in a patient portal. Surg Endosc 2016 Apr;30(4):1432-1440 [FREE Full text] [doi: 10.1007/s00464-015-4347-y] [Medline: 26123340]
32. Schickedanz A, Huang D, Lopez A, Cheung E, Lyles CR, Bodenheimer T, et al. Access, interest, and attitudes toward electronic communication for health care among patients in the medical safety net. J Gen Intern Med 2013 Jul;28(7):914-920 [FREE Full text] [doi: 10.1007/s11606-012-2329-5] [Medline: 23423453]
33. Ancker JS, Barrón Y, Rockoff ML, Hauser D, Pichardo M, Szerencsy A, et al. Use of an electronic patient portal among disadvantaged populations. J Gen Intern Med 2011 Oct;26(10):1117-1123 [FREE Full text] [doi: 10.1007/s11606-011-1749-y] [Medline: 21647748]
34. Thomas SB, Fine MJ, Ibrahim SA. Health disparities: The importance of culture and health communication. Am J Public Health 2004 Dec;94(12):2050. [doi: 10.2105/AJPH.94.12.2050]
35. van Ryn M. Research on the provider contribution to race/ethnicity disparities in medical care. Med Care 2002;40(1):I140-I151. [doi: 10.1097/00005650-200201001-00015]

36. van Ryn M, Fu SS. Paved with good intentions: do public health and human service providers contribute to racial/ethnic disparities in health? Am J Public Health 2003 Feb;93(2):248-255. [Medline: 12554578]
37. Demiris G, Afrin LB, Speedie S, Courtney KL, Sondhi M, Vimarlund V, et al. Patient-centered applications: use of information technology to promote disease management and wellness. A white paper by the AMIA knowledge in motion working group. J Am Med Inform Assoc 2008 Feb;15(1):8-13 [FREE Full text] [doi: 10.1197/jamia.M2492] [Medline: 17947617]
38. Ford EW, Hesse BW, Huerta TR. Personal health record use in the United States: forecasting future adoption levels. J Med Internet Res 2016 Mar 30;18(3):e73 [FREE Full text] [doi: 10.2196/jmir.4973] [Medline: 27030105]
39. Lustria ML, Smith SA, Hinnant CC. Exploring digital divides: an examination of eHealth technology use in health information seeking, communication and personal health information management in the USA. Health Informatics J 2011 Sep;17(3):224-243. [doi: 10.1177/1460458211414843] [Medline: 21937464]
40. Pew Internet. 2005. How Women and Men Use the Internet URL: http://www.pewinternet.org/2005/12/28/how-women-and-men-use-the-internet/ [accessed 2017-08-25] [WebCite Cache ID 6pe5mvwap]
41. Graetz I, Gordon N, Fung V, Hamity C, Reed ME. The digital divide and patient portals: Internet access explained differences in patient portal use for secure messaging by age, race, and income. Med Care 2016 Aug;54(8):772-779. [doi: 10.1097/MLR.0000000000000560] [Medline: 27314262]
42. Brooks RG, Menachemi N. Physicians' use of email with patients: factors influencing electronic communication and adherence to best practices. J Med Internet Res 2006;8(1):e2 [FREE Full text] [doi: 10.2196/jmir.8.1.e2] [Medline: 16585026]
43. Wagner EH, Aiello Bowles EJ, Greene SM, Tuzzio L, Wiese CJ, Kirlin B, et al. The quality of cancer patient experience: perspectives of patients, family members, providers and experts. Qual Saf Health Care 2010 Dec;19(6):484-489. [doi: 10.1136/qshc.2010.042374] [Medline: 21127109]
44. Mazor KM, Beard RL, Alexander GL, Arora NK, Firneno C, Gaglio B, et al. Patients' and family members' views on patient-centered communication during cancer care. Psychooncology 2013 Nov;22(11):2487-2495 [FREE Full text] [doi: 10.1002/pon.3317] [Medline: 23780672]
45. Rao S, Brammer C, McKethan A, Buntin MB. Health information technology: transforming chronic disease management and care transitions. Prim Care 2012 Jun;39(2):327-344. [doi: 10.1016/j.pop.2012.03.006] [Medline: 22608869]
46. Groves RM. Nonresponse rates and nonresponse bias in household surveys. Public Opin 2006;70(5):646-675 [FREE Full text] [doi: 10.1093/poq/nfl033]

Abbreviations

CATI: Computer-assisted telephone interviewing
HINTS: Health Information National Trends Survey
OR: odds ratio
PHR: personal health record
USPS: United States Postal Service

Edited by G Eysenbach; submitted 24.04.17; peer-reviewed by J Tavares, D Blanch-Hartigan, CF Chung, L Garvin, P Shafer, S Tuo; comments to author 12.07.17; revised version received 03.10.17; accepted 16.11.17; published 29.03.18.

Please cite as:
Tarver WL, Menser T, Hesse BW, Johnson TJ, Beckjord E, Ford EW, Huerta TR
Growth Dynamics of Patient-Provider Internet Communication: Trend Analysis Using the Health Information National Trends Survey (2003 to 2013)
J Med Internet Res 2018;20(3):e109
URL: http://www.jmir.org/2018/3/e109/
doi:10.2196/jmir.7851
PMID:29599107

Original Paper

Detecting Motor Impairment in Early Parkinson's Disease via Natural Typing Interaction With Keyboards: Validation of the neuroQWERTY Approach in an Uncontrolled At-Home Setting

Teresa Arroyo-Gallego[1,2,3,4], MSc; María J Ledesma-Carbayo[2,3], PhD; Ian Butterworth[5], MSc; Michele Matarazzo[6,7,8,9], MD; Paloma Montero-Escribano[10], MD; Verónica Puertas-Martín[7], PhD; Martha L Gray[1,5], PhD; Luca Giancardo[5,11*], PhD; Álvaro Sánchez-Ferro[5,6,7,8,12*], MD, PhD

[1]Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, United States

[2]Biomedical Image Technologies, Universidad Politécnica de Madrid, Madrid, Spain

[3]Biomedical Research Networking Centre thematic area of Bioengineering, Biomaterials and Nanomedicine, Madrid, Spain

[4]nQ Medical Inc, Cambridge, MA, United States

[5]Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, United States

[6]Centro Integral de Neurociencias A.C., Hospital Universitario HM Puerta del Sur, Móstoles, Spain

[7]Neurology Department, Instituto de Investigación del Hospital 12 de Octubre, Madrid, Spain

[8]Enfermedades Neurodegenerativas, Centro de Investigación Biomédica en Red, Madrid, Spain

[9]Pacific Parkinson's Research Centre, The University of British Columbia, Vancouver, BC, Canada

[10]Movement Disorders Unit, Hospital Clínico San Carlos, Madrid, Spain

[11]Center for Precision Health, School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX, United States

[12]Medical School, CEU-San Pablo University, Madrid, Spain

[*]these authors contributed equally

Corresponding Author:
Teresa Arroyo-Gallego, MSc
Institute for Medical Engineering and Science
Massachusetts Institute of Technology
E25-413
77 Massachusetts Avenue
Cambridge, MA, 02139
United States
Phone: 1 857 523 6281
Email: tarroyog@mit.edu

Abstract

Background: Parkinson's disease (PD) is the second most prevalent neurodegenerative disease and one of the most common forms of movement disorder. Although there is no known cure for PD, existing therapies can provide effective symptomatic relief. However, optimal titration is crucial to avoid adverse effects. Today, decision making for PD management is challenging because it relies on subjective clinical evaluations that require a visit to the clinic. This challenge has motivated recent research initiatives to develop tools that can be used by nonspecialists to assess psychomotor impairment. Among these emerging solutions, we recently reported the neuroQWERTY index, a new digital marker able to detect motor impairment in an early PD cohort through the analysis of the key press and release timing data collected during a controlled in-clinic typing task.

Objective: The aim of this study was to extend the in-clinic implementation to an at-home implementation by validating the applicability of the neuroQWERTY approach in an uncontrolled at-home setting, using the typing data from subjects' natural interaction with their laptop to enable remote and unobtrusive assessment of PD signs.

Methods: We implemented the data-collection platform and software to enable access and storage of the typing data generated by users while using their computer at home. We recruited a total of 60 participants; of these participants 52 (25 people with Parkinson's and 27 healthy controls) provided enough data to complete the analysis. Finally, to evaluate whether our in-clinic-built

algorithm could be used in an uncontrolled at-home setting, we compared its performance on the data collected during the controlled typing task in the clinic and the results of our method using the data passively collected at home.

Results: Despite the randomness and sparsity introduced by the uncontrolled setting, our algorithm performed nearly as well in the at-home data (area under the receiver operating characteristic curve [AUC] of 0.76 and sensitivity/specificity of 0.73/0.69) as it did when used to evaluate the in-clinic data (AUC 0.83 and sensitivity/specificity of 0.77/0.72). Moreover, the keystroke metrics presented a strong correlation between the 2 typing settings, which suggests a minimal influence of the in-clinic typing task in users' normal typing.

Conclusions: The finding that an algorithm trained on data from an in-clinic setting has comparable performance with that tested on data collected through naturalistic at-home computer use reinforces the hypothesis that subtle differences in motor function can be detected from typing behavior. This work represents another step toward an objective, user-convenient, and quasi-continuous monitoring tool for PD.

(J Med Internet Res 2018;20(3):e89) doi:10.2196/jmir.9462

KEYWORDS

eHealth; machine learning; telemedicine

Introduction

Background

Parkinson's disease (PD) is the second most prevalent neurodegenerative disorder affecting 0.3% of the general population and about 1% in people over 60 years [1]. Today, PD diagnosis and management rely on the clinical judgment of neurologists to detect and evaluate the severity of motor and nonmotor manifestations of the disease [2]. The Unified Parkinson's Disease Rating Scale (UPDRS) is the most widely used method to assess the longitudinal course of PD [3,4]. The UPDRS score comprises 4 subscales, including the clinician-scored motor section (UPDRS-III) that provides a comprehensive evaluation of PD motor degeneration through the evaluation of the observed performance in a series of specific motor tasks [5]. Administered typically by a movement disorder specialist, this scale requires significant training to minimize rater bias [6]. The need of a trained specialist intrinsically limits the frequency at which disease status and progression can be assessed to a number of on-site clinical evaluations, usually every 2 to 6 months [7].

There has been substantial interest in the last decade to develop tools that can assess motor function in PD without the need for specialist training or even on-site administration [8]. Such tools could complement the current standard, introducing the potential for greater screening opportunities or an increased assessment frequency for tracking changes. A variety of technological approaches have been designed for use in the clinic, such as finger-tapping that introduces a series of standardized finger-movement tasks that provide quantitative measurements of motor impairment [9]. Additionally, out-of-clinic approaches have been trialed, such as the mPower initiative [10], a smartphone-based activity tracker that collects longitudinal data from a series of tasks and surveys specifically designed to evaluate the progression of PD symptoms.

Objective

Our project focuses on the analysis of finger-keyboard interaction to assess psychomotor impairment. We have previously shown that we can extract information relevant to users' psychomotor status by timing the keystroke events during a typing task using a mechanical keyboard [11]. In Giancardo et al [12], we showed that it was possible to derive an early-PD phenotype based on a metric derived from the typing data acquired in a controlled clinical environment. Subjects were asked to transcribe a randomly selected folktale using a word processor on a standard 15-inch laptop during a 15-min timed routine.

The widespread use of personal electronics has placed typing among the activities of our daily routine. This enables the possibility of leveraging the data from users' natural interaction with their devices to apply our method in an unobtrusive manner. From a data-collection standpoint, it is straightforward to extend our technology to collect timing information in a naturalistic ecologically valid scenario (eg, home). However, from the standpoint of data analysis, passive monitoring poses interesting challenges that could affect the application of our method to evaluate at-home natural typing. From a data-sampling perspective, typing happens in unpredictable bursts that introduce a high degree of sparsity in the resulting typing signals. The various contexts in which the typing data are generated at-home may also add difficulty in contrast with the controlled copy task performed in the clinic. Finally, hardware heterogeneity introduces a potential confounder in the at-home setting, which we were able to control in our in-clinic setting using a single machine approach.

In this paper, we present the results of the validation of our in-clinic-built algorithm to detect PD typing patterns in an uncontrolled at-home setting. We implemented a data-collection platform that allowed us to passively collect the typing information from subjects' daily interaction with their laptop. Our algorithm performed well prospectively on a controlled typing study conducted in the clinic. Here, we examine whether the same algorithm performs well on the typing data collected at home and evaluate the influence of the in-clinic typing task in subjects' normal typing behaviors.

Methods

Study

The results presented in this work analyze the baseline data collected as part of a 6-month longitudinal PD

drug-responsiveness study (NCT02522065). All the experimental protocols were approved by the Massachusetts Institute of Technology, USA (Committee on the Use of Humans as Experimental Subjects approval no. 1412006804), HM Hospitales, Spain (No. 15.05.796-GHM), Hospital 12 de Octubre, Spain (No. CEIC:14/090), and Hospital Clínico San Carlos, Spain (No.14/136-E). All subjects provided informed consent before study enrollment. The recruitment and experimental procedures were carried out following the relevant institutional guidelines.

The study cohort consisted of 60 subjects, 30 people with recently diagnosed Parkinson's (PwP) and 30 healthy controls. Only subjects who self-reported at least 30 min of daily laptop use were considered for the study. The exclusion criteria included cognitive impairment, upper limb functional limitation, sleep disorders, and use of antipsychotics or sedative drugs. At the moment of enrollment, 6 PwP were on rasagiline while the remaining 24 were completely drug naïve. Notably, unlike levodopa or dopamine agonists, rasagiline is a compound that has a little impact on motor performance; so, for the purposes of this study related to motor performance, we considered patients on rasagiline to be similar to the PwP who had not yet started medication. They maintained their baseline medication status (ie, no drug or continued rasagiline) for a period of time after the enrollment visit.

The enrolled participants underwent an initial baseline assessment in the clinic that included clinical evaluation, an in-clinic controlled typing test, and the technical setup to enable at-home monitoring. The medical examination included a UPDRS-III-based evaluation carried out by movement disorder specialists. For the in-clinic typing test, the participants were asked to transcribe an unstandardized sample text on a standard word processor during 15 min. To emulate natural interaction with the device, subjects were asked to type as they would normally do at home. A standard machine was used in the in-clinic setting, specifically, a Lenovo G50-70 i3-4005U with 4GB of memory and a 15-inch screen running Manjaro Linux operative system. While undertaking the test, the data-collection software ran in the background. Once the task was completed, the typing data were sent to our database server. As part of the baseline visit workflow, the data-collection software was installed on participants' personal laptop to enable at-home remote monitoring. If they shared their computer, we provided them with a laptop with preinstalled software. Subjects were encouraged to enter into the routine of typing an email or a document for at least 15 min per day but otherwise use the computer as they would do normally.

Once enrolled in the study, PwP subjects kept their baseline medication status for about a week. This baseline period allowed an unbiased comparison between the in-clinic and at-home conditions on the assessment of our method. Due to the naturalistic design, there was some variability in the time between the initial visit and the date the new therapy was started (ie, some variability in the duration of the baseline period). This period ranged from 0 to 63 days. For the data reported here, we used a 7-day baseline period, unless there was a medication change within that timeframe, in which case we used the actual baseline period. For the control group, the baseline period was defined as the 7-day period since the date they first logged in to the neuroQWERTY platform.

To assure a comparable amount of typing activity between the in-clinic and at-home settings, only subjects who aggregated at least 15 min of typing data during their corresponding at-home baseline period were included in the analysis. Though, importantly, data at home were sparsely distributed over the multi-day baseline period, whereas the in-clinic data were concentrated in a 15-min continuous typing task. To manage this sparsity in the at-home data, we applied the concept of valid window to filter typing gaps and low-activity intervals. A valid window was defined as a data sequence of at least 30 keystrokes within a 90-s time interval. We excluded 5 PwP and 3 control subjects from the analysis because they did not reach the equivalent 15-min active typing threshold (10 valid windows) during the baseline period.

A summary of demographic and clinical information for the resulting cohort, 25 PwP and 27 healthy controls, can be found in Table 1. Regarding PD severity, all PwP subjects were newly diagnosed cases and in the very early stages of the disease, with a mean UPDRS-III score of 20.48 points. For reference, a score of 20 points is typical of patients with very mild disease severity [13]. The 2 groups were matched in age, gender, and volume of daily typing. A detailed representation of the at-home baseline data collected for each subject is shown in Figure 1. The plot illustrates the heterogeneity of subjects' typing behaviors, which we previously identified as one of the potential risks for the validation of our approach in a natural at-home setting. Participants typed an average of 24.07 (SD 15.13) min per day, with 2.79 min per day for the less-active subject and 83.14 min per day for the most active subject. This variability was also observed within subjects' typing routines, as several participants did not present a consistent typing activity over the monitored time period. These characteristics in the at-home spontaneous typing data contrast with the quasi-continuous signal captured during the in-clinic typing test. In the *Analysis* subsection, we will explain how we addressed these differences to allow us to compare the performance of the algorithm for the in-clinic and at-home scenarios.

Table 1. Comparison of the clinical and demographic variables between the Parkinson's disease and control groups. From the total participants, 52 provided a sufficient amount of at-home typing data (a cumulative total of at least 15 min). The UPDRS-III scale ranges from 0 to 108 (a higher score indicates more severe impairment and disability). For reference, a score of 20 points is typical of patients with very mild disease severity.

Variable	People with Parkinson's (n=25)	Healthy controls (n=27)	*P* value
UPDRS-III[a], mean (SD)	20.48 (6.56)	1.93 (1.84)	<.001
Age in years, mean (SD)	60.2 (12.0)	60.81 (10.63)	.73
Number of women, n (%)	12 (48)	14 (52)	.79
Number of men, n (%)	13 (52)	13 (48)	.79
Daily typing in minutes, mean (SD)	24.58 (15.91)	23.58 (14.68)	.61

[a]UPDRS-III: Unified Parkinson's Disease Rating Scale (Part III).

Data-Collection Platform

The neuroQWERTY platform provides functionality for user registration and log-in, distribution of the data-collection software, and storage and management of the typing data. Once installed, the data-collection software runs in the background, capturing the timing information of any keyboard input. More specifically, for each keystroke, the program stores the timestamps corresponding to the press and release events. To ensure privacy, the collected information did not include the content of each specific key. However, each keystroke was labeled with its corresponding key category; special key, right side key, or left side key, to allow filtering of key types that engage nonstandard digit kinematics (eg, SHIFT). The mean for measured temporal resolution of the data-collection software was 3 (SD 0.28) msec.

The typing information, linked to each user account, was automatically sent to a remote server for analysis. Privacy and data security were assured at 3 levels: at the client level, the data transmission level, and the data storage level. Any typing data stored on the local machine (which again, did not include the content of the keys) were encrypted and deleted from the device after sending to the remote server. Data transmission was protected through secure hypertext transfer protocol. At the server level, data were stored in the database in an encrypted format and were only accessible by authorized database administrators or by the user himself after authentication.

Finally, the platform included an administrator module to provide the study coordinators with an interface to access and control participants' typing activity. The administrator dashboard implemented a color code to alert study coordinators about users' prolonged inactivity. Web-based visualization of the subjects' typing data was also enabled, including the daily key count and the temporal representation of the raw key typing dynamics. A schema of the complete neuroQWERTY platform framework is shown in Figure 2.

Analysis

We evaluate the classification ability of the neuroQWERTY index (nQi) to separate a group of healthy controls from an early PD population using the typing data collected during subjects' natural interaction with their laptop. The nQi is the output of a computational algorithm that uses the information contained in the sequences of hold times, the time between pressing and releasing each key on a mechanical keyboard, to detect evidence of PD motor impairment. This algorithm was first introduced in Giancardo et al [12], where we showed its ability to accurately discriminate early PwP from healthy controls by analyzing the data collected in a controlled in-clinic typing task.

A representation of the algorithm pipeline is shown in Figure 3. The hold time signal is split into 90-s windows that are analyzed as independent typing units. Applying variance analysis, the information within each unit is reduced to a 7D feature vector that is used as the input of an ensemble model consisting of a family of linear support vector regressors (SVR). An independent window-level score is calculated as the median of the outputs of each linear SVR. Finally, the final nQi score is computed as the mean of the window-level scores. The feature analysis and algorithm parameter estimation are described in detail in our previous paper, Giancardo et al [12].

The ensemble linear-SVR model was trained using an external dataset that included the typing signals of 18 early-PD subjects and 13 healthy controls different than the ones included in this study. This training set was fully collected in a controlled in-clinic environment, that is, during a timed copy task integrated into the study clinical visit. Therefore, the main question we try to answer in this work is whether our algorithm can generalize to typing data acquired in a fully uncontrolled home-based scenario where the subjects are free to use their laptops as they normally do.

Figure 1. At-home typing activity. Panel A represents the amount of typing data collected from each of the 52 subjects (25 PwP, 27 CNT) included in the analysis. The red (PwP) and blue (CNT) color scales indicate daily typing activity measured as the number of valid typing windows provided by each subject during the analysis period. We defined a valid window as a sequence of at least 30 keystrokes within 90 s. Panel B illustrates the variability in the amount of typing data with an example from a single PwP subject.

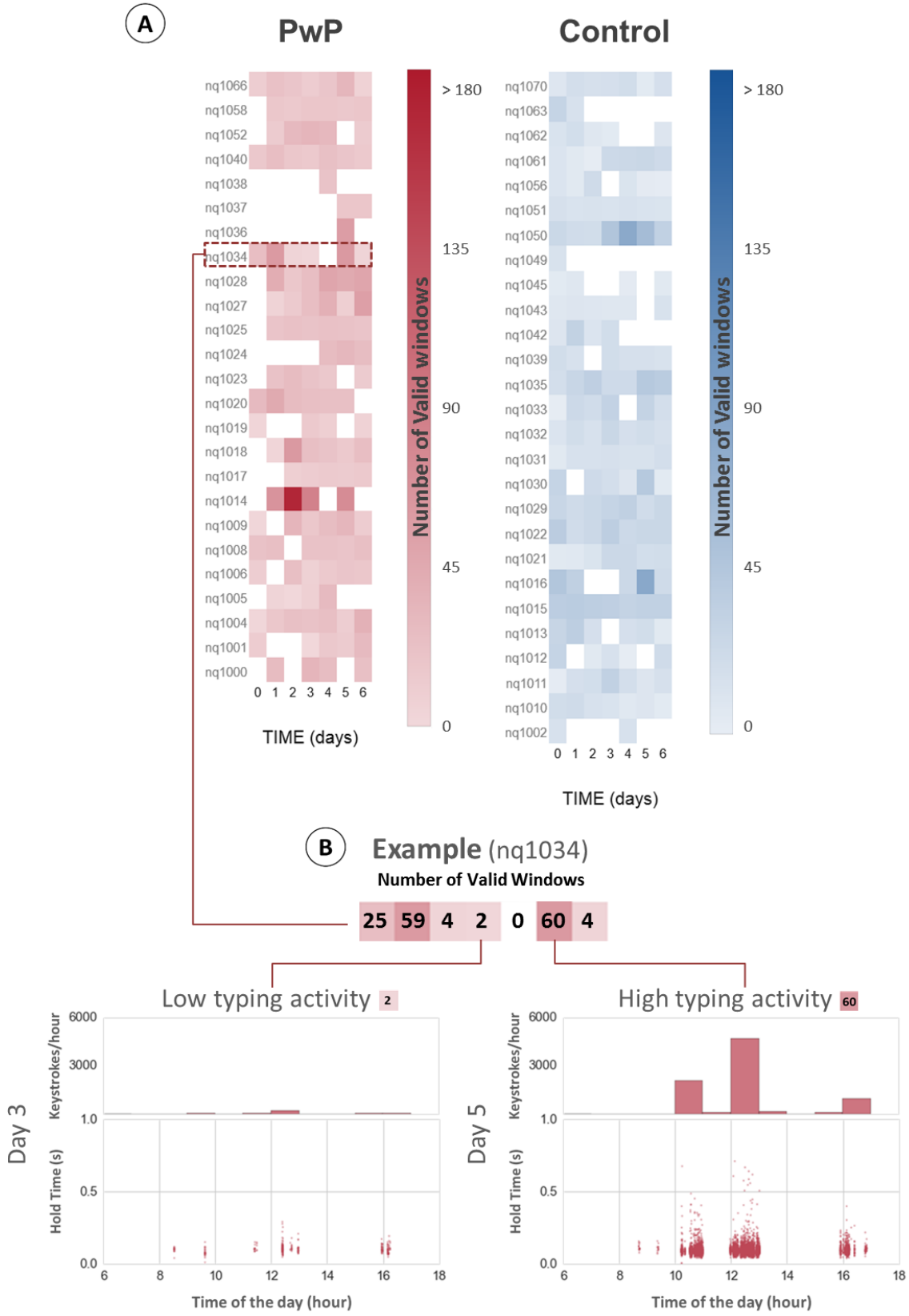


Figure 2. The neuroQWERTY platform. This platform was designed to allow for automatic data retrieval of typing data collected at home and remote management by a study coordinator. Operationally, an account in the neuroQWERTY platform was created for each participant in the study. The data-collection software was downloaded and installed in their users' personal laptop to enable remote data collection. The data, linked to each user account, was encrypted and automatically sent to a remote server through their home Internet connection. The neuroQWERTY platform also implemented an administrator module to provide the study coordinators with an interface to control and visualize participants' typing activity.

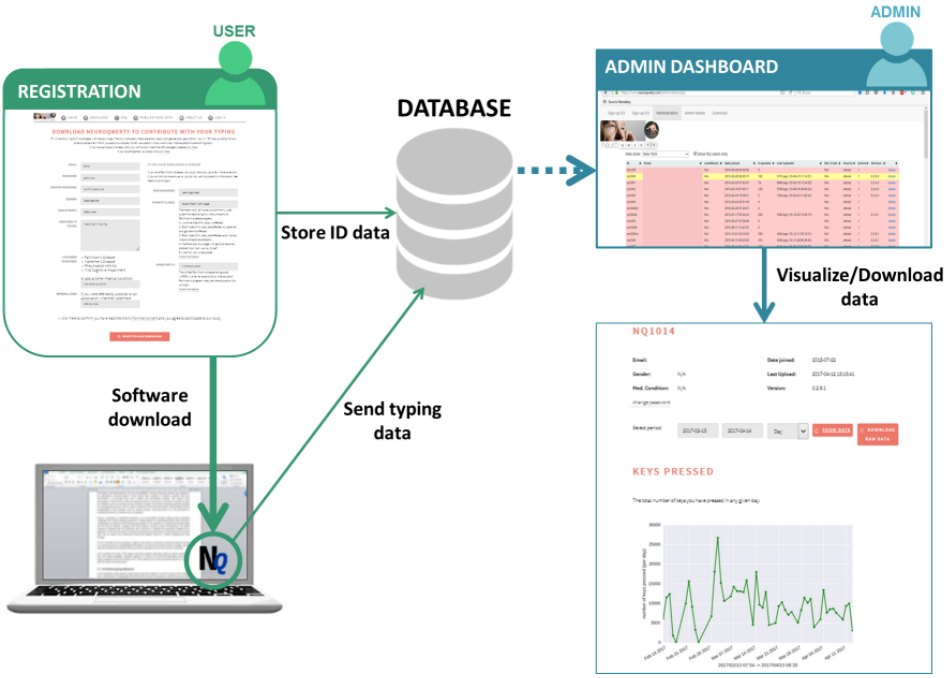
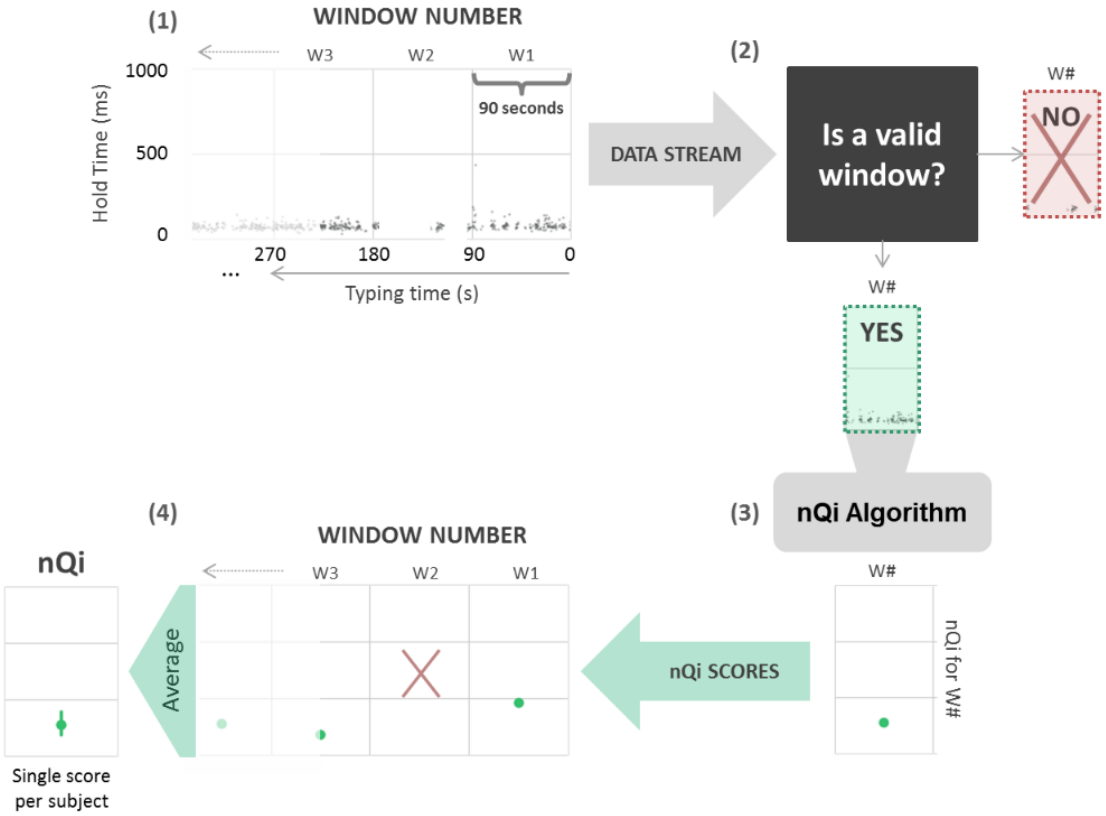


Figure 3. Algorithm pipeline. The figure represents the pipeline to generate a single neuroQWERTY index (nQi) from a stream of typing data. (1) The typing signal is defined as the time series of hold times corresponding to each keystroke within a typing routine. This signal is split by nonoverlapping 90-s windows that the algorithm will evaluate as independent typing units. (2) Only windows with at least 30 keystrokes within the 90-s interval are analyzed. (3) The neuroQWERTY algorithm, previously trained on a separate in-clinic dataset, computes a single numerical score from each independent window. (4) The final nQi is computed as the average of the window-level scores.



In Figures 4 and 5, we present an example of the application of the algorithm in a controlled in-clinic task opposed to the same process in an at-home typing setting. The comparison of the hold time data representation between in-clinic (panel A) and at-home (panel B) illustrates the sparsity introduced in the typing signals by the uncontrolled at-home environment, where the keyboard is only used intermittently as opposed to the continuous stream of data collected in the controlled in-clinic setting. To account for the sparsity of the hold time series, only valid windows, subsequences of at least 30 keystrokes within each 90-s interval, are included in the analysis. Special key types (eg, SHIFT) that may engage nonstandard digit kinematics are excluded from the hold time (HT) data collection. Due to the duration of the at-home baseline period, the volume of data collected at home is generally greater than the in-clinic data available for each subject. We measured an average of 9.62 (2.13) valid windows per subject during the in-clinic typing test and collected an average of 112.33 (70.65) valid windows per subject from the 7-day at-home typing activity. The extended at-home monitoring period increases subjects' at-home sample size, which tends to reduce the individual's internal variance intensified in this uncontrolled environment.

The data analysis comprised 2 phases. First, we evaluated the influence of the controlled typing task in subjects' normal typing behaviors. Specifically, we compared the measured values of the raw typing metrics, flight time (FT, delay between consecutive key presses), and HT (time between pressing and releasing a key), and the computed nQi scores between the in-clinic and at-home typing settings. To assess the similarity in the relationship between the in-clinic and at-home metrics, we computed the line of best fit and correlation coefficient. We completed this first part of the analysis with a Bland-Altman plot [14] to evaluate the nQi score's agreement between the 2 typing settings.

In the second part of the analysis, we assessed the classification performance of the neuroQWERTY method using the at-home typing data and compared these results with the ones obtained in the clinic. The results obtained in each typing settings were evaluated using the following metrics: receiver operating characteristic (ROC) analysis and the Mann-Whitney *U* test to reject the null hypothesis that the healthy controls and the Parkinson's samples come from the same distribution. For the ROC analysis, we used a sampling with replacement method to define a distribution of curves from which we computed the average area under the ROC curve (AUC) and its CIs. Each curve is built on an iterative process that monotonically increases the value of the index to define a dynamic threshold. On each iteration, a sensitivity/specificity pair is computed using the current threshold value. These pairs are used to draw the resulting ROC curve. The value of the AUC can be interpreted as the probability of the classifier to rank a randomly chosen positive instance higher than a randomly chosen negative one [15]. To evaluate the equivalence of our method between the in-clinic and at-home settings, we estimated the percentage agreement and the statistical difference of the resulting ROC curves (DeLong test [15]).

Figure 4. Example of the application of the neuroQWERTY algorithm in an in-clinic typing test.

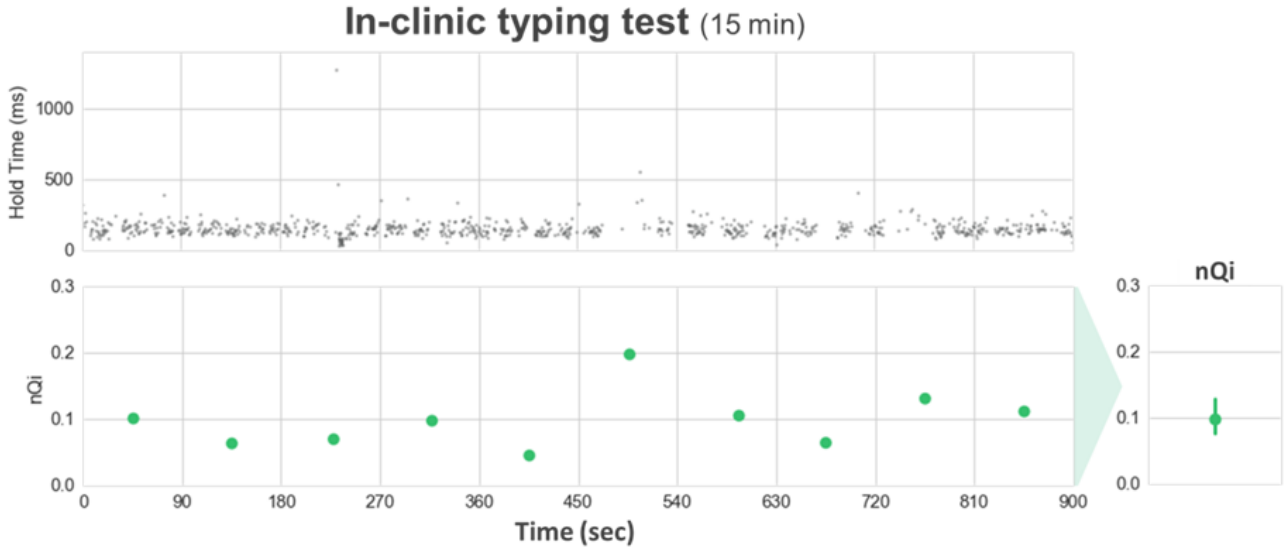
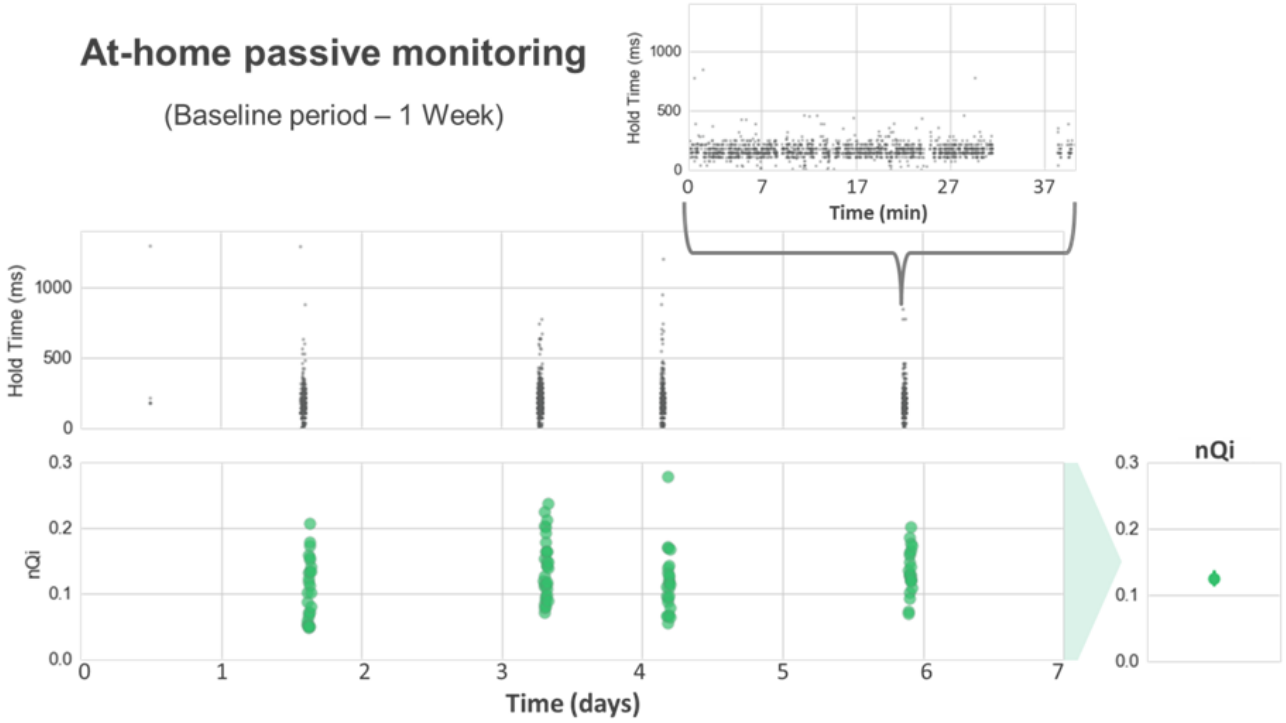


Figure 5. Example of the application of the neuroQWERTY algorithm in the at-home setting. The neuroQWERTY algorithm described in Figure 3 can be used indistinctly to evaluate controlled or natural typing data. This figure represents the at-home typing data and corresponding scores for the same subject shown in Figure 4 (note different time scales used in Figure 4 and Figure 5). Although the uncontrolled activity appears in unpredictable bursts that introduce a high degree of sparsity, our window-based approach allows to analyze the at-home data using the same method applied for the quasi-continuous in-clinic data.



Results

The results of the raw typing variables agreement between in-clinic and at-home are shown in Figure 6. We evaluate the statistical relationship, line of best fit, and correlation for the median flight and HT measured in-clinic and at-home settings. The values of the 2 typing metrics are very similar independently of the typing scenario, as shown by correlation coefficient values, .913 for the median FT and .897 for the median HT, and also by the slope of the computed line of best fit, close to 1 in both cases.

A similar analysis applied to the nQi scores is shown in Figure 7. The linearity between the in-clinic and at-home settings for this variable is weaker than that observed on the raw typing variables. The correlation coefficient is .749 in this case, and the slope of the line of best fit is not as close to the unit, .597. However, the agreement analysis suggests a correspondence between the scores measured in-clinic and at-home settings, with a 92% (48/52) of the cases falling between the Bland-Altman limits of agreement (LoA).

In terms of classification performance, the nQi worked well with the at-home typing data (Figure 8 and Table 2). The absolute nQi scores tended to be larger for at-home data relative to the corresponding in-clinic values (Figures 7 and 8), but in both cases the scores for PwP were generally greater than for healthy controls. The similarity in classification performance for in-clinic versus at-home data can also be seen by comparing the ROC curves (Figure 8 and Table 2). The cutoff point was estimated using the closest-to-(0,1), that is, the use case that maximizes the sensitivity/specificity pair [16] (Table 2). The neuroQWERTY algorithm discriminates our early PD population from healthy controls with an AUC of 0.76 (0.66-0.88) using the typing data from the at-home natural interaction on a mechanical keyboard. In the clinic, the results of the analysis, a controlled typing task in the same cohort, achieved an AUC of 0.83 (0.74-0.92). According to the DeLong test, the ROC AUC difference between in-clinic and at-home settings was not significant (P=.18). The percentage agreement of the results of our method between the 2 typing settings was 79%.

Figure 6. Comparison of raw typing metrics between in-clinic and at-home typing settings. The figure shows the correlation of the raw typing metrics, hold time (HT; time between pressing and releasing a key), and flight time (FT; delay between two consecutive key presses), between in-clinic and at-home settings. Each point represents the metric coordinates (in-clinic, at-home) for each of the 52 participants included in the analysis. Both HT and FT values are very similar independently of the typing scenario, as shown by the correlation coefficient values. These results suggest that the in-clinic task does not alter the way subjects type in comparison with their natural typing at-home, which supports our hypothesis that the neuroQWERTY algorithm, built in an in-clinic setting, could be applied to evaluate motor impairment using the typing data from an uncontrolled at-home setting.

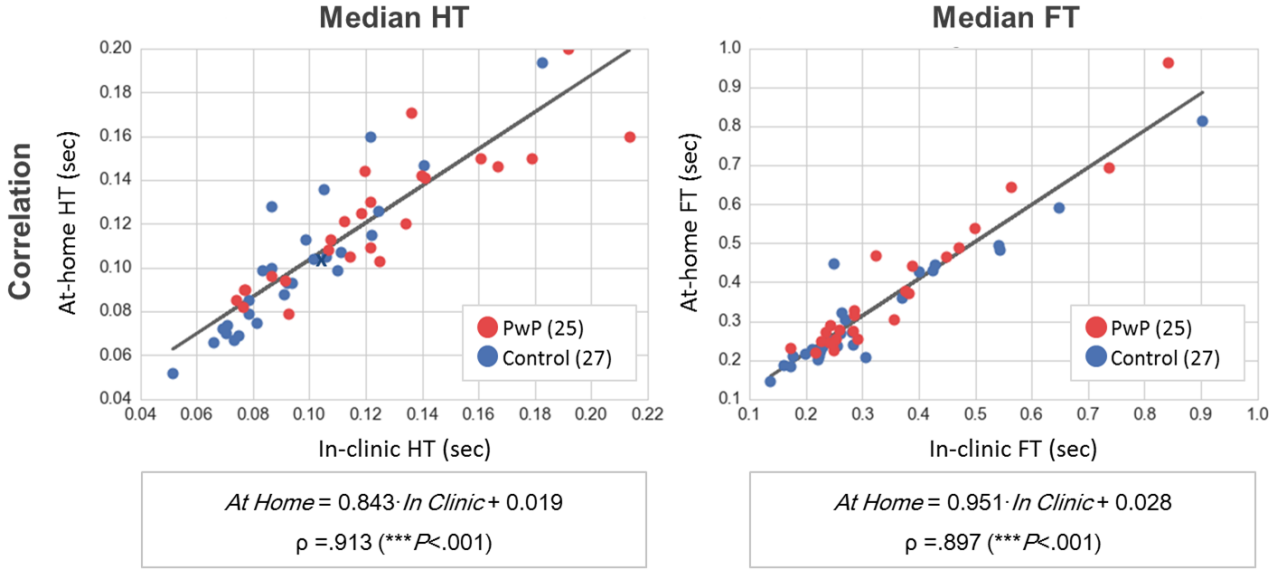


Figure 7. Comparison of neuroQWERTY index (nQi) between in-clinic and at-home typing settings. We evaluated the influence of the typing setting in the nQi scores by applying a similar analysis as described in Figure 6 for the raw typing metrics. Panel A shows the correlation of the nQi scores computed in-clinic and at-home. Panel B includes the results of the Bland-Altman analysis to evaluate the agreement of our method in the two typing scenarios. The black line shows the mean difference (d) and the top and bottom dashed lines show the limits of agreement (LoA, d±1.96×SDd).

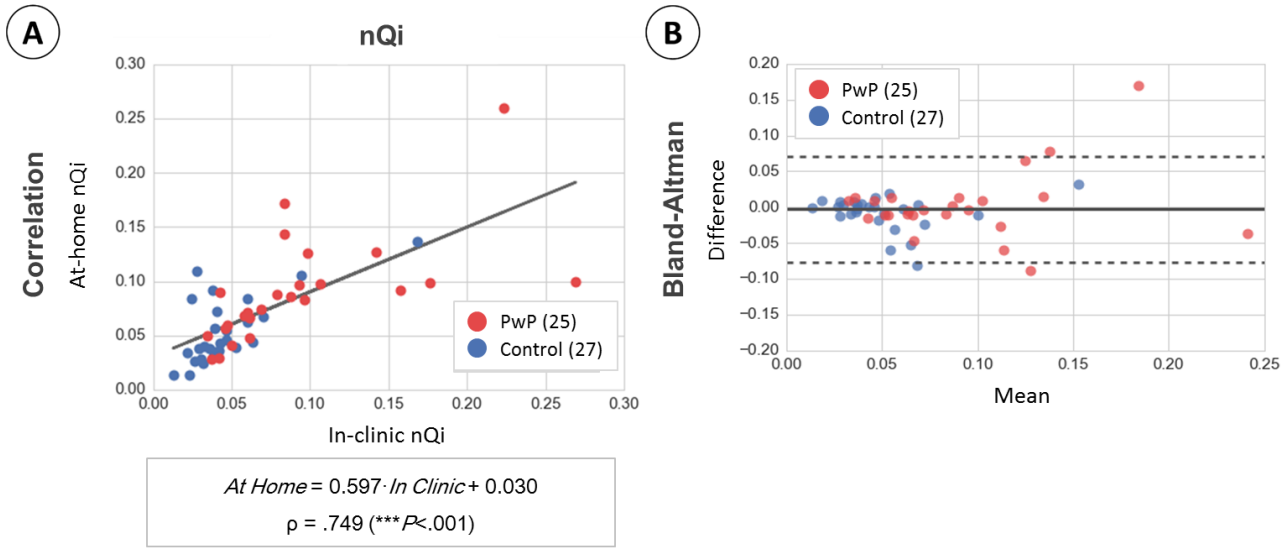


Figure 8. Comparison of neuroQWERTY index (nQi) performance between in-clinic and at-home typing settings. Panel A scatterplot illustrates the in-clinic and at-home nQi scores in a patient level. The two black lines represent the classification thresholds computed in-clinic (nQi=0.0473) and at-home (nQi=0.0667). These thresholds were estimated for closest-to-(0,1) cutoff points that maximize sensitivity/specificity pairs. Panel B presents the comparison of the receiver operating characteristic (ROC) curves showing the classification rate for the in-clinic and at-home nQi. The plotted curves are the average result of the bootstrapped ROC analysis and the shadowed areas represent the corresponding CIs [5th-95th]. The statistical significance of the Mann-Whitney U test is estimated to reject the null hypothesis that the two groups, PwP and CNT, come from the same population. It is noted as: P<.001(***), P<.01(**), and P<.05(*).

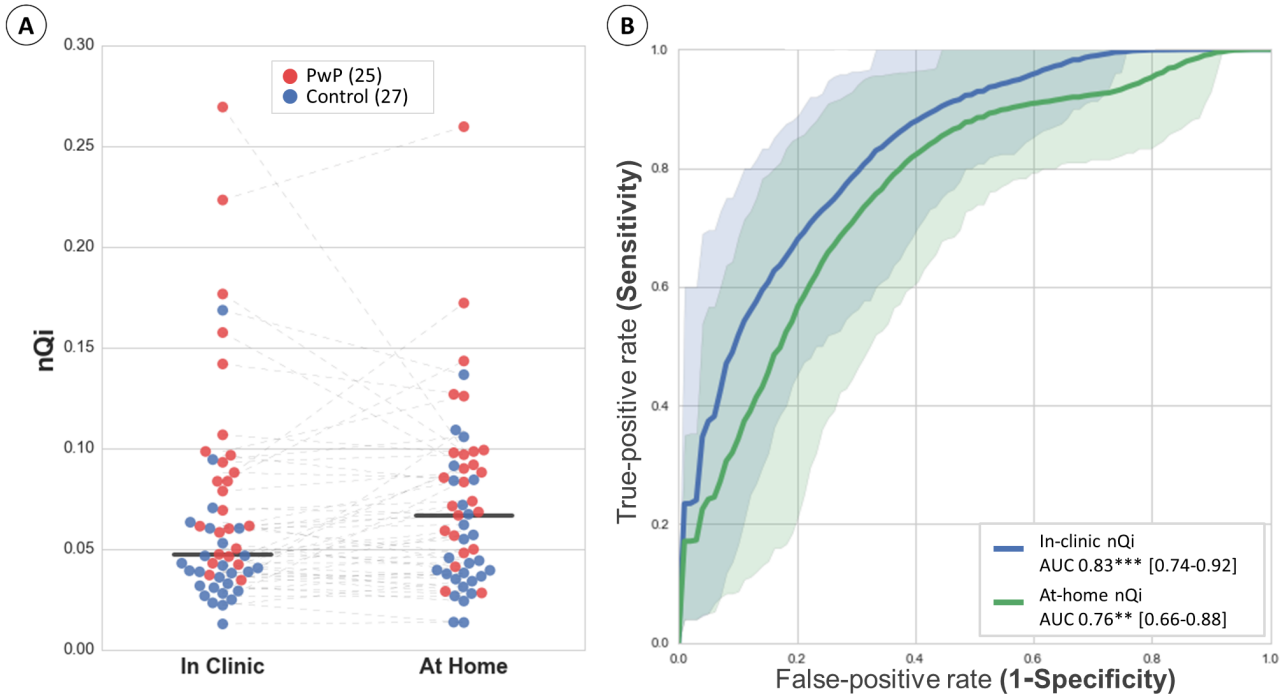


Table 2. The neuroQWERTY index (nQi) performance comparison. The classification performance achieved at-Home is comparable with the results obtained in a controlled in-clinic. The statistical significance is computed with 2-sided Mann-Whitney *U* test to reject the null hypothesis that PwP and healthy control subjects come from the same population.

Metric	nQi[a] Score	
	In-clinic	At-home
Mean (SD) for PwP[b] (n=25)	0.092 (0.058)	0.090 (0.048)
Mean (SD) for healthy controls (n=27)	0.046 (0.029)	0.054 (0.030)
AUC[c] (5th-95th)	0.83 (0.74-0.92)	0.76 (0.66-0.88)
Significance	*P*<.001	*P*<.01
Sensitivity/specificity	0.77/0.72	0.73/0.69
DeLong test	*P*=.18	*P*=.18
Percentage agreement	79%	79%

[a]nQi: neuroQWERTY index.

[b]PwP: people with Parkinson's.

[c]AUC: area under the curve.

Discussion

Principal Findings

The results of this study represent a step toward a transparent and ubiquitous motor sign assessment tool for PD. In our previous work [12], we introduced the neuroQWERTY method, a machine learning algorithm trained to quantify PD severity through the analysis of the typing patterns found in the time series of HT. Our method was able to discriminate an early PD population from a matched control group using the typing data collected during a controlled in-clinic task. In this paper, we tested the validity of our algorithm in an uncontrolled at-home setting. The neuroQWERTY platform allowed us to unobtrusively collect the typing information from a cohort comprising 30 PwP and 30 matched healthy controls. Most (52/60, 90%) of the study subjects (25 PwP, 27 healthy controls) provided enough data during the follow-up period to evaluate the nQi at home. Our neuroQWERTY algorithm, built using a separate in-clinic dataset, was able to distinguish PwP from

healthy controls through the analysis of natural at-home typing patterns with an AUC of 0.76 and 0.73/0.69 sensitivity/specificity. Despite the sparsity and heterogeneity introduced by each subject's routine use of the computer, the neuroQWERTY method performed nearly as well in the at-home setting as it did when applied in a controlled in-clinic typing task (AUC 0.83 and 0.77/0.72 sensitivity/specificity). The nQi scores presented no significant differences between the de-novo PwP group (19) and the subset of PwP participants on medication (Multimedia Appendix 1).

The comparison of the raw typing metrics between the 2 typing scenarios suggests that the in-clinic typing test does not affect the way people type with regard to their normal use of the computer at-home. The correlation coefficient for the median HT between in-clinic and at-home was ρ=.913 (P<.001). A similar analysis applied to the resulting nQi shows a weaker correlation between the scores computed from the in-clinic and at-home typing data (ρ=.749, P<.001, with 48 out of 52 or 92% of the samples within the Bland-Altman LoA). This could be due to the sensitivity of the algorithm to small changes in the HT values between the 2 typing settings. Despite the weaker correlation, the classification performance of the neuroQWERTY method applied at-home was similar to the classification performance in-clinic (statistically indistinguishable by the DeLong test: P=.18, percentage agreement: 79%).

These results support our initial hypothesis that PD-related motor signs affect the way patients interact with mechanical keyboards and are, therefore, detectable through the analysis of their regular typing patterns. The ability of the neuroQWERTY algorithm to extrapolate the patterns learned from a separate in-clinic dataset to correctly identify PD-characteristics in the at-home typing data provides external validity to our method. Being able to generalize to data collected from the hardest possible scenario can also be seen as an opportunity to improve these results by implementing an at-home-specific algorithm that, trained on passively collected data, will be able to identify the useful information and learn to filter the several different sources of noise introduced by the uncontrolled at-home setting. Although the current clinical standard, UPDRS, outperforms our technique, the goal of neuroQWERTY is not to replace UPDRS but to provide a method that enables PD assessment when a clinician is not available. Nevertheless, it would be interesting to explore the potential for the nQi approach to provide a meaningful indication of UPDRS. Despite being based just on a distal upper limb movement, with our limited dataset we did find a significant moderate correlation between the nQi scores and UPDRS-III (Multimedia Appendix 2).

Main Contribution and Limitations

Using the timing information from users' natural typing activity provides our approach with a number of advantages over alternative solutions, but it also poses some limitations. An obvious concern is the level of compliance, since the method depends on sufficient use of the computer. In the study cohort, a high percentage of the participants (90%) provided enough data during the 7-day follow-up to shape a representative typing pattern. Our user adherence results highlight the advantages of passive data collection in contrast with other existing active task-based methods. Task-based methods are commonly limited by their dependence on users' active engagement to collect information through a series of standardized tasks, which introduces potential artifacts due to subject-awareness of being monitored [17] and hinders user compliance. As an example, in the context of the mPower study less than 10% of the participants provided 5 or more finger-tapping data points over a 6-month follow-up period [10].

Although our passive data-collection approach significantly increases user adherence, some strategies could be employed to maximize it. A possible solution to reduce the rate of excluded participants would be collecting data not only from laptop use but from any electronic device that entails typing. In Arroyo et al [18], it was proven that a similar approach can be used to detect PD via smartphone touchscreen typing. Integrating data from multiple devices would provide a more continuous stream of data; therefore, a deeper insight to assess PD signs.

Future Work

Proving that our method can distinguish an early PwP cohort, with an average years from diagnosis of 1.66 (1.20) and mean UPDRS-III score of 20.48 (6.56) from a matched healthy control group is an indicator that at-home typing patterns can capture PD-specific motor characteristics that are mild in this stage of the disease. This could have an impact in early detection of PD as machine learning algorithms can be trained to detect very subtle variations in the input data, in this case changes in the typing patterns, caused by early motor manifestations of PD that may often go unnoticed by clinicians [19]. The neuroQWERTY software could be installed on PD-risk populations' devices to enable earlier diagnosis, when putative neuroprotective treatments could stop neurodegeneration. Clinical studies in an as-yet-undiagnosed population would be needed to validate the sensitivity and applicability of our tool for this specific use case.

Although our classification results show promise, our longer term goal is to develop a tool to objectively track progression of PD signs. This would provide clinicians with invaluable information to tailor treatments to patients' specific conditions. Today, there is no known cure for PD, but available medications can help manage its symptoms. Individualized treatment regimens are crucial to provide optimized symptom control [20]. Medications adjustments rely mainly on the information gathered by movement disorder experts during clinical visits. This limits decision making to subjective follow-up examinations scheduled every 2 to 6 months. Ideally, our approach could be applied not only to classify but also to track PD progression and therapeutic efficacy. This would require further validation in a longitudinal study to evaluate the precision of the neuroQWERTY approach to monitor PD progression over time.

Conclusions

Relying on the analysis of the temporal patterns from the daily interaction with electronic devices, our approach introduces a new way to objectively and unobtrusively detect motor impairment in PD, providing access to quasi-continuous

ambulatory data without harming user compliance. The main purpose of this analysis was to evaluate the validity of the nQi, an in-clinic-built digital marker for early PD motor impairment, in an uncontrolled at-home setting. The classification performance of the algorithm was statistically similar in its ability to discriminate 25 PwP and 27 healthy controls from the at-home typing data (AUC of 0.76 and 0.73/0.69 sensitivity/specificity) nearly as well as it was able to separate them using the in-clinic typing patterns (AUC 0.83 and 0.77/0.72 sensitivity/specificity). These results prove that the data collected from subjects' routine use of the computer are also valid to detect PD-related motor signs, getting us closer to our ultimate goal of providing an objective ambulatory tool to monitor PD progression.

Acknowledgments

This research is being financially supported by the Comunidad de Madrid, Fundación Ramón Areces, and The Michael J Fox Foundation for Parkinson's research (grant number 10860). We thank the M+Vision and MIT LinQ faculty for their guidance in developing this project. We also thank our many clinical collaborators at MGH in Boston, Hospital 12 de Octubre, Hospital Clínico and Centro Integral en Neurociencias HM CINAC in Madrid for their insightful contributions. The research leading to these results has received funding from Consejería de Educación, Juventud y Deporte of Comunidad de Madrid and the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement no. 291820.

Conflicts of Interest

TAG is currently employed by nQ Medical Inc, a company that has licensed the neuroQWERTY technology.

Multimedia Appendix 1

Influence of medication in nQi assessment. Six of the PwP study participants were on rasagiline when they joined the study. Panel A shows no significant differences between the nQi scores of the PwP participants using rasagiline and the de-novo group. Panel B shows the UPDRS-III scores for the CNT, PwP De-Novo and PwP On-Medication groups. Participants on rasagiline scored higher up in the motor scale in comparison with the average score in the full PwP cohort. It is possible that this greater severity in their baseline status could be masking the effect of medication.

[PNG File, 330KB - jmir_v20i3e89_app1.png]

Multimedia Appendix 2

nQi correlation with motor clinical standard. The figure shows the correlation between UPDRS-III scale and the nQi scores measured in-clinic (Panel A) and at-home (Panel B). Correlations were significant in both typing settings and moderate as shown by the correlation coefficients; .50 in-clinic and .34 at-home. Despite being based just on distal upper limb movement, with our limited dataset we did find a significant moderate correlation between the nQi scores and UPDRS-III.

[PNG File, 160KB - jmir_v20i3e89_app2.png]

References

1. de Lau M, Breteler M. Epidemiology of Parkinson's disease. Lancet Neurol 2006;5(6):525-535. [Medline: 16713924]
2. Gelb DJ, Oliver E, Gilman S. Diagnostic criteria for Parkinson disease. Arch Neurol 1999 Jan;56(1):33-39. [Medline: 9923759]
3. Ramaker C, Marinus J, Stiggelbout A, van HB. Systematic evaluation of rating scales for impairment and disability in Parkinson?s disease. Mov Disord 2002;17(5):867-876. [Medline: 12360535]
4. Whone A, Watts R, Stoessl A, Davis M, Reske S, Nahmias C, REAL-PET Study Group. Slower progression of Parkinson's disease with ropinirole versus levodopa: the REAL-PET study. Ann Neurol 2003 Jul;54(1):93-101. [doi: 10.1002/ana.10609] [Medline: 12838524]
5. Martínez-Martín P, Gil-Nagel A, Gracia LM, Gómez JB, Martínez-Sarriés J, Bermejo F. Unified Parkinson's disease rating scale characteristics and structure. The cooperative multicentric group. Mov Disord 1994 Jan;9(1):76-83. [doi: 10.1002/mds.870090112] [Medline: 8139608]
6. Taylor TA, Jefferis G, Koop M, Hill BC, Hastie T, Heit G, et al. Quantitative measurements of alternating finger tapping in Parkinson's disease correlate with UPDRS motor disability and reveal the improvement in fine motor control from medication and deep brain stimulation. Mov Disord 2005 Oct;20(10):1286-1298. [doi: 10.1002/mds.20556] [Medline: 16001401]
7. von Campenhausen S, Winter Y, Rodrigues ES, Sampaio C, Ruzicka E, Barone P, et al. Costs of illness and care in Parkinson's disease: an evaluation in six countries. Eur Neuropsychopharmacol 2011 Feb;21(2):180-191. [doi: 10.1016/j.euroneuro.2010.08.002] [Medline: 20888737]

8. Sánchez-Ferro Á, Elshehabi M, Godinho C, Salkovic D, Hobert M, Domingos J, et al. New methods for the assessment of Parkinson's disease (2005 to 2015): a systematic review. Mov Disord 2016 Sep;31(9):1283-1292. [doi: 10.1002/mds.26723] [Medline: 27430969]
9. Shimoyama I, Ninchoji T, Uemura K. The finger-tapping test. A quantitative analysis. Arch Neurol 1990 Jun;47(6):681-684. [Medline: 2346396]
10. Bot B, Suver C, Neto E, Kellen M, Klein A, Bare C, et al. The mPower study, Parkinson disease mobile data collected using ResearchKit. Sci Data 2016 Mar 03;3:160011 [FREE Full text] [doi: 10.1038/sdata.2016.11] [Medline: 26938265]
11. Giancardo L, Sánchez-Ferro A, Butterworth I, Mendoza CS, Hooker JM. Psychomotor impairment detection via finger interactions with a computer keyboard during natural typing. Sci Rep 2015 Apr 16;5:9678 [FREE Full text] [doi: 10.1038/srep09678] [Medline: 25882641]
12. Giancardo L, Sánchez-Ferro A, Arroyo-Gallego T, Butterworth I, Mendoza CS, Montero P, et al. Computer keyboard interaction as an indicator of early Parkinson's disease. Sci Rep 2016 Oct 05;6:34468 [FREE Full text] [doi: 10.1038/srep34468] [Medline: 27703257]
13. Martínez-Martín P, Rodríguez-Blázquez C, Mario A, Arakaki T, Arillo VC, Chaná P, et al. Parkinson's disease severity levels and MDS-unified Parkinson's disease rating scale. Parkinsonism Relat Disord 2015 Jan;21(1):50-54. [doi: 10.1016/j.parkreldis.2014.10.026] [Medline: 25466406]
14. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986 Feb 8;1(8476):307-310. [Medline: 2868172]
15. Fawcett T. An introduction to ROC analysis. Pattern Recognit Lett 2006 Jun;27(8):861-874. [doi: 10.1016/j.patrec.2005.10.010]
16. Perkins NJ, Schisterman EF. The inconsistency of “optimal” cutpoints obtained using two criteria based on the receiver operating characteristic curve. Am J Epidemiol 2006 Apr 01;163(7):670-675 [FREE Full text] [doi: 10.1093/aje/kwj063] [Medline: 16410346]
17. McCarney R, Warner J, Iliffe S, van HR, Griffin M, Fisher P. The hawthorne effect: a randomised, controlled trial. BMC Med Res Methodol 2007;7:30 [FREE Full text] [doi: 10.1186/1471-2288-7-30] [Medline: 17608932]
18. Arroyo-Gallego T, Ledesma-Carbayo M, Sanchez-Ferro A, Butterworth I, Mendoza CS, Matarazzo M, et al. Detection of motor impairment in Parkinson's disease via mobile touchscreen typing. IEEE Trans Biomed Eng 2017 Sep;64(9):1994-2002. [doi: 10.1109/TBME.2017.2664802] [Medline: 28237917]
19. Ross GW, Abbott RD, Petrovitch H, Tanner CM, White LR. Pre-motor features of Parkinson's disease: the Honolulu-Asia aging study experience. Parkinsonism Relat Disord 2012 Jan;18 Suppl 1:S199-S202. [doi: 10.1016/S1353-8020(11)70062-1] [Medline: 22166434]
20. Tarakad A, Jankovic J. Diagnosis and management of Parkinson's disease. Semin Neurol 2017 Apr;37(2):118-126. [doi: 10.1055/s-0037-1601888] [Medline: 28511252]

Abbreviations

AUC: area under the ROC curve
HT: hold time
LoA: limits of agreement
nQi: neuroQWERTY index
PD: Parkinson’s disease
PwP: People with Parkinson’s
ROC: receiver operating characteristic
SVR: support vector regressor
UPDRS-III: Unified Parkinson's disease rating scale, Part III

Edited by G Eysenbach; submitted 20.11.17; peer-reviewed by D Newman, C Fernández; comments to author 13.12.17; revised version received 24.12.17; accepted 24.12.17; published 26.03.18.

Please cite as:
Arroyo-Gallego T, Ledesma-Carbayo MJ, Butterworth I, Matarazzo M, Montero-Escribano P, Puertas-Martín V, Gray ML, Giancardo L, Sánchez-Ferro Á
Detecting Motor Impairment in Early Parkinson’s Disease via Natural Typing Interaction With Keyboards: Validation of the neuroQWERTY Approach in an Uncontrolled At-Home Setting
J Med Internet Res 2018;20(3):e89
URL: http://www.jmir.org/2018/3/e89/
doi:10.2196/jmir.9462
PMID:29581092

Short Paper

Effect of Harm Anchors in Visual Displays of Test Results on Patient Perceptions of Urgency About Near-Normal Values: Experimental Study

Brian J Zikmund-Fisher[1,2,3], PhD; Aaron M Scherer[3,4], PhD; Holly O Witteman[5,6,7,8], PhD; Jacob B Solomon[3], PhD; Nicole L Exe[3], MPH; Angela Fagerlin[9,10], PhD

[1]Department of Health Behavior and Health Education, University of Michigan, Ann Arbor, MI, United States
[2]Department of Internal Medicine, University of Michigan, Ann Arbor, MI, United States
[3]Center for Bioethics and Social Sciences in Medicine, University of Michigan, Ann Arbor, MI, United States
[4]Department of Internal Medicine, University of Iowa, Iowa City, IA, United States
[5]Department of Family and Emergency Medicine, Laval University, Quebec City, QC, Canada
[6]Office of Education and Professional Development, Faculty of Medicine, Laval University, Quebec City, QC, Canada
[7]Population Health and Optimal Health Practices Research Unit, Research Center of the CHU de Québec-Université Laval, Quebec City, QC, Canada
[8]Laval University Research Institute for Primary Care and Health Services, Quebec City, QC, Canada
[9]Department of Population Health Sciences, University of Utah, Salt Lake City, UT, United States
[10]Salt Lake City Veterans Affairs Center for Informatics Decision Enhancement and Surveillance, Salt Lake City, UT, United States

Corresponding Author:
Brian J Zikmund-Fisher, PhD
Department of Health Behavior and Health Education
University of Michigan
1415 Washington Heights
Ann Arbor, MI, 48109-2029
United States
Phone: 1 7349369179
Email: bzikmund@umich.edu

Related Article:

This is a corrected version. See correction statement: https://www.jmir.org/2025/1/e74908

Abstract

Background: Patient-facing displays of laboratory test results typically provide patients with one reference point (the “standard range”).

Objective: To test the effect of including an additional harm anchor reference point in visual displays of laboratory test results, which indicates how far outside of the standard range values would need to be in order to suggest substantial patient risk.

Methods: Using a demographically diverse, online sample, we compared the reactions of 1618 adults in the United States who viewed visual line displays that included both standard range and harm anchor reference points (“Many doctors are not concerned until here”) to displays that included either (1) only a standard range, (2) standard range plus evaluative categories (eg, “borderline high”), or (3) a color gradient showing degree of deviation from the standard range.

Results: Providing the harm anchor reference point significantly reduced perceived urgency of close-to-normal alanine aminotransferase and creatinine results (*P* values <.001) but not generally for platelet count results. Notably, display type did not significantly alter perceptions of more extreme results in potentially harmful ranges. Harm anchors also substantially reduced the number of participants who wanted to contact their doctor urgently or go to the hospital about these test results.

Conclusions: Presenting patients with evaluative cues regarding when test results become clinically concerning can reduce the perceived urgency of out-of-range results that do not require immediate clinical action.

(J Med Internet Res 2018;20(3):e98) doi:10.2196/jmir.8889

KEYWORDS

decision making; education of patients; electronic health record; computer graphics; clinical laboratory information systems

Introduction

Patients can increasingly view their laboratory test results directly via patient portals of electronic health record (EHR) systems [1]. While patients value such information to enable self-management and support informed patient-provider interactions [2-5], access to test results does not guarantee that patients can understand or use that information to improve their health or their care [6]. In particular, most current EHR patient portals present test results to patients in tables [7], which is a format that is difficult for many patients to interpret, especially those with lower numeracy or literacy skills [8]. In addition, EHR portals typically only provide patients with one reference point (the standard range) to aid in interpreting such data. Patients who receive out-of-range test values may have little idea how alarmed they should, or should not, feel [5,6,9,10]. Consequently, patients may call their doctor for an urgent appointment or even go to the hospital for something that would be more appropriately managed through regular follow-up visits [10].

In order to improve patient understanding of laboratory test results, our research team used user-centered design principles to develop several visual number line formats for presenting these types of data [11]. In an experimental test, displaying test results in these number line formats instead of tables increased user sensitivity to test result variations [12]. We also showed that many people interpret all results outside of the standard range as equally urgent, even though many slightly out-of-range results are not, in fact, clinically concerning.

The core problem that patients face is one of information evaluability [13,14]. Many patients lack training and experience with most laboratory tests, so they cannot necessarily map a particular test result to its meaning (ie, how good or bad it is, or how much risk it represents). As long as most patients lack meaningful reference points beyond the standard range, they will struggle to discriminate between different types of non-normal test results.

Here, we present results from additional data collected at the same time as the previous study in which we tested a visual display format that added a second reference point to indicate how far outside of the standard range a test value needs to be to become clinically concerning. Our objective was to determine whether providing such "harm anchors" would reduce the perceived urgency of near-normal values without significantly altering perceptions of extreme values, thereby increasing overall sensitivity to test result variations.

Methods

Participants were a demographically diverse, stratified random sample of adults in the United States recruited during May 2016 from a panel of Internet users, and the survey was administered by Survey Sampling International. The design, sampling process, data management procedures, and outcome measures received exempt status approval from the University of Michigan Health Sciences and Behavioral Sciences Institutional Review Board.

Participants were asked to imagine that they were using an online EHR portal to view laboratory test results and then saw three specific test results (platelet count, alanine aminotransferase [ALT], and serum creatinine). Each test result was initially shown as slightly outside of the standard range and then as a more extreme test result. All participants viewed a platelet count of 135 x 10^9/L and then 25 x 10^9/L, an ALT value of 80 U/L and then 360 U/L, and a creatinine value of 2.2 mg/dL and then 3.4 mg/dL.

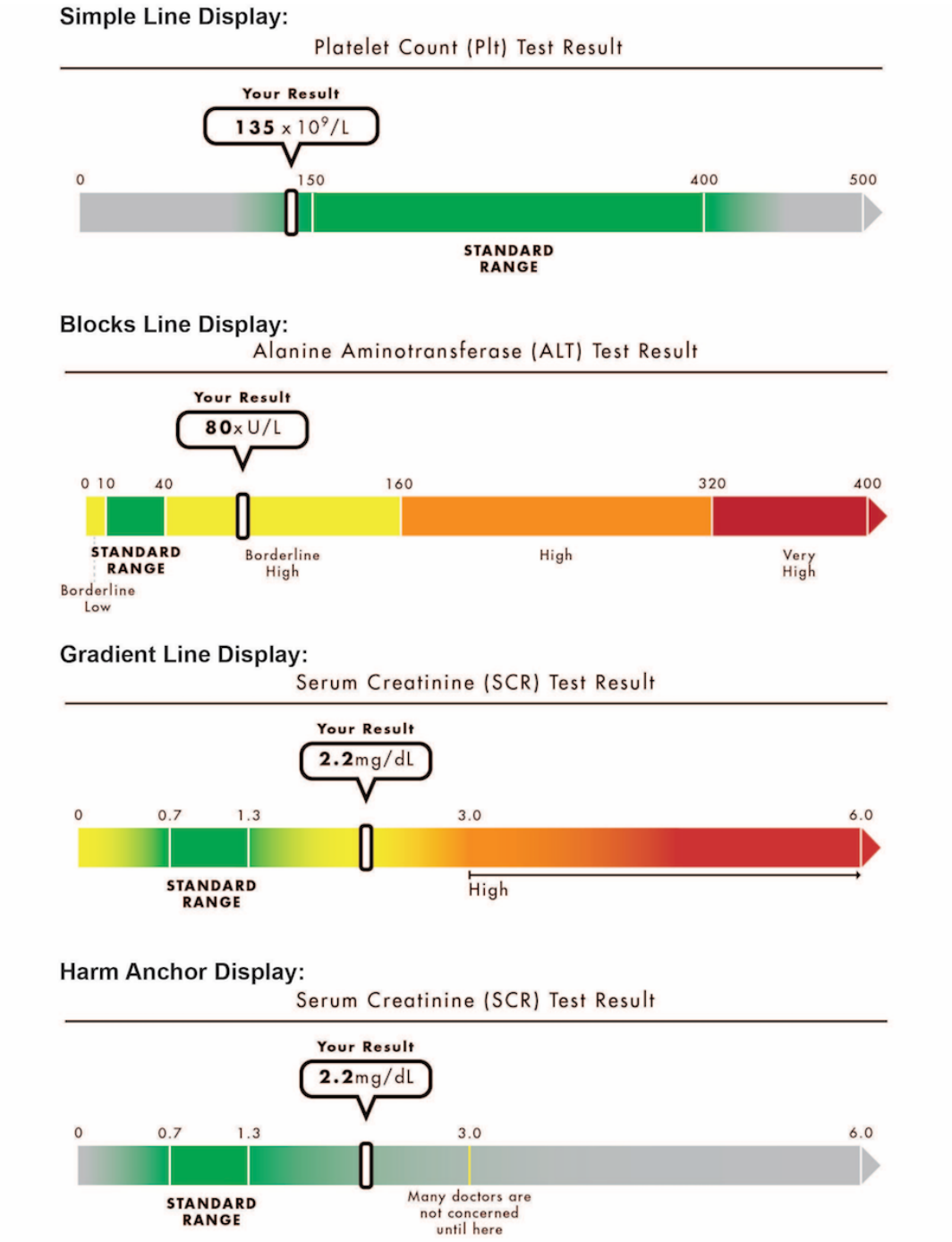
In a previous paper, we compared participant reactions to three visual number line graph displays (Figure 1): simple line displays that show the standard range and endpoints, but have no other visual reference points; blocks line displays that included color-coded evaluative categories (eg, "borderline high"); and gradient line displays that used a color gradient to indicate the extent of deviation from the standard range [12]. The range of values shown and the evaluative labels and categories displayed were selected as plausible values based on input from several clinician members of our research team.

In this analysis, we compared the results from those three previously published conditions to data collected at the same time from an additional (randomized) group of participants. These participants received displays that included an added harm anchor (ie, a threshold line outside of the standard range labeled, "many doctors are not concerned until here") but were otherwise identical to the simple line displays (see Figure 1, bottom image). This language was developed through several iterations of pretesting with patients. Based on consultations with multiple clinician collaborators, we selected the anchor levels shown (platelets=100 x 10^9/L; ALT=160 U/L; creatinine=3.0 mg/dL) as plausible approximations of the point at which nonnormal values require more urgent attention for most patients.

The primary outcome measure was respondents' subjective sense of urgency to the displayed test results. We averaged respondents' responses to two questions: "How alarming does this result feel to you?" and, "How urgent of an issue is this result?" (both measured on 6-point Likert scales; 1=*not at all*, 6=*very*). The resulting measure of subjective urgency showed high reliability for all tests and test results (Cronbach alpha=.91 to .95). We measured behavioral intentions by asking whether respondents would initiate a new contact with a health professional (eg, by calling their doctor for an urgent appointment or going to a hospital) versus either waiting until their next regular appointment or doing nothing (for question details, see Zikmund-Fisher, et al [12]). To measure display format preferences, we used respondents' average responses on a set of four questions (Cronbach alpha=.87) that asked how well the images described the results, how helpful they were, whether respondents would trust the images, and whether respondents would like to see results presented in these formats (all measured using 5-point Likert scales) [12]. We report

one-way analyses of variance (ANOVAs) with Bonferroni corrections to compare ratings of perceived urgency and user preferences, and chi-squared tests to compare willingness to wait. All analyses were performed using STATA 14 [15]. All tests of significance were two-sided and used alpha=.05.

Figure 1. Examples of each of the four visual display formats and three tests included in this study.



Results

A total of 1618 adult participants (aged 19 to 89 years) completed the survey and were randomized to the experimental conditions examined here. See Table 1 for respondent characteristics.

As shown in Table 2, providing the harm anchor reference point on the visual display significantly reduced perceived urgency of the close-to-normal ALT and creatinine results compared to all three other display formats (all P values <.001) without significantly altering perceptions of more extreme results in potentially harmful ranges. For platelet count results, however, we observed only a mildly significant difference between the harm anchor display and the blocks display. Use of the harm anchor labels also substantially reduced the number of participants who wanted to contact their doctor urgently or go to the hospital when shown near-normal ALT or creatinine test results (Table 3).

Overall, there were no significant differences in participants' preferences among the four display types (Harm Anchor mean=3.77 vs Simple Line mean=3.62, Blocks Line mean=3.76, Gradient Line mean=3.68; ANOVA F(3, 1615)=2.33, *P*=.07).

Table 1. Respondent characteristics.

Characteristic	Value
Age, mean (range)	48.8 (19,89)
Gender, n (%)	
Male	769 (47.6)
Female	842 (52.1)
Transgender	4 (0.25)
Race, n (%)	
White	1255 (77.8)
Black	213 (13.2)
Asian	64 (4.0)
Native American	13 (0.8)
Other / Multi-race	69 (4.3)
Hispanic	212 (13.2)
Bachelor's degree or higher education, n (%)	810 (50.1)
Current health, n (%)	
Excellent	248 (15.3)
Very good	605 (37.4)
Good	522 (32.3)
Fair	213 (13.2)
Poor	29 (1.8)

Table 2. Perceived urgency of near-normal and more extreme test results, by display type. Perceived urgency was measured on a 1-6 scale, with higher numbers corresponding to greater perceived urgency. *P* values were calculated by post-hoc comparisons following one-way analyses of variance with Bonferroni corrections for multiple comparisons. ALT: alanine aminotransferase.

Test result	Harm Anchor	Simple Line		Blocks Line		Gradient Line	
		Rating	*P* value[a]	Rating	*P* value[a]	Rating	*P* value[a]
Near-normal results							
Platelets=135 x 10^9/L	3.66	3.72	>.99	3.94	.016	3.73	>.99
ALT=80 U/L	3.08	4.00	<.001	3.96	<.001	3.56	<.001
Creatinine=2.2 mg/dL	3.52	4.09	<.001	3.99	<.001	3.91	<.001
Extreme results							
Platelets=25 x 10^9/L	5.09	5.26	.32	5.20	>.99	5.30	.10
ALT=360 U/L	5.32	5.44	.78	5.35	>.99	5.39	.10
Creatinine=3.4 mg/dL	4.71	4.81	>.99	4.58	.48	4.74	>.99

[a]*P* values reported are for comparisons to the harm anchor condition.

Table 3. Percentage of participants reporting intentions to contact their doctor urgently or go to the hospital based on their near-normal test results, by display type. ALT: alanine aminotransferase.

Test result	Harm Anchor	Simple Line	Blocks Line	Gradient Line	Overall test	
					Chi square statistic	*P* value
Platelets=135 x 10^9/L	44.2%	50.0%	53.3%	51.6%	$\chi^2(3)=7.61$	.06
ALT=80 U/L	34.7%	55.8%	58.1%	48.2%	$\chi^2(3)=53.83$	<.001
Creatinine=2.2 mg/dL	35.2%	56.7%	53.5%	52.3%	$\chi^2(3)=45.15$	<.001

Discussion

As hypothesized, presenting patients with cues regarding the values at which particular test results become clinically concerning reduced respondents' perceptions of urgency about certain types of out-of-range results that were not of immediate clinical concern. Our results suggest that including harm anchors in test result communications could, in certain circumstances, provide important benefits to patients by increasing the evaluability of variations among out-of-range results.

As in our previous comparison of visual line displays versus tabular displays [12], the effect size we observed varied substantially across the three tests presented. A possible explanation for this finding is that it is a function of the relative size of the standard reference range, as compared to the range of values shown. The largest effect of harm anchors was observed for ALT tests, which have a narrow reference range within a large range of possible values. By contrast, we observed minimal to no effect of harm anchors on displays of platelet counts; a test for which deviations of 50 to 100 (x 10^9/L) outside of the standard range (only 20%-40% of the width of the standard range) represent significant changes in patient risk. Thus, harm anchors may be most useful when communicating with patients about unfamiliar tests that can have wide ranges of potential variation.

Operationalization of this idea, however, will require overcoming several challenges. First, harm anchors, by definition, represent clinical judgment, and different clinicians may reasonably disagree regarding the point at which harm threshold should be set [16]. Second, even if harm thresholds could be agreed upon, the point at which a patient (or clinician) should view a test result as requiring urgent action should logically vary based on patient characteristics or medical context (eg, initial diagnosis vs long term management). We also acknowledge the primary limitations of this study: the use of a hypothetical scenario and the testing of the harm anchor concept within the constraints of a particular visual display design. Inclusion of harm anchor information within other types of visual displays or tables might result in different findings than those observed here.

Nonetheless, our results demonstrate that designing displays to inform patients regarding what is dangerous, as opposed to what is considered usual or normal, might offer practical benefits. We suggest that harm anchors or other risk-related reference points should be considered when designing patient-facing displays of health data in order to increase the interpretability of such communications. These types of displays should be most useful in situations where relatively unfamiliar laboratory tests are being conducted for monitoring purposes (eg, monitoring liver or kidney function while on extended medication regimens). These situations are likely to result in mild deviations in test result values that, while important to monitor, are not immediately concerning to clinicians. Enabling patients to know that these mild deviations are not urgent will reduce patient worry and might also minimize unnecessary patient requests for urgent appointments when routine follow-up would be sufficient.

Acknowledgments

This work was previously presented at the Annual Meeting of the Society for Medical Decision Making, Vancouver, BC, Canada; October 24, 2016.

Funding for this research was provided entirely by a grant from the United States Agency for Healthcare Research and Quality to BJZ-F (R01 HS021681). The funding agreement assured the authors' independence in: designing the study; the collection, analysis, and reporting of the data; and the decision to submit the article for publication. HOW receives salary support from a Research Scholar Junior 1 Career Development Award by the Fonds de Recherche du Québec-Santé.

The authors acknowledge the assistance of Sandeep Vijan, MD and Kenneth Langa, MD in determining the appropriate display ranges and categorization schemas for each of the types of test results used in this study. In addition, we acknowledge the graphic design efforts of Patrick Winkel in creating the images used in this study, as well as the work of Grace Bienek in developing earlier versions of these visual designs. Lastly, we are grateful for the guidance and inspiration provided by the patient members of our research team: Margaret Newton, Stephanie Burke, and James Piazza.

Conflicts of Interest

None declared.

References

1. American Hospital Association. Washington, DC: American Hospital Association; 2016. Individuals' ability to electronically access their hospital medical records, perform key tasks is growing URL: http://www.aha.org/research/reports/tw/16jul-tw-healthIT.pdf [accessed 2016-10-29] [WebCite Cache ID 6t5QyAjYU]
2. Cimino JJ, Patel VL, Kushniruk AW. The patient clinical information system (PatCIS): technical solutions for and experience with giving patients access to their electronic medical records. Int J Med Inform 2002 Dec 18;68(1-3):113-127. [Medline: 12467796]
3. Keselman A, Slaughter L, Smith CA, Kim H, Divita G, Browne A, et al. Towards consumer-friendly PHRs: patients' experience with reviewing their health records. AMIA Annu Symp Proc 2007:399-403 [FREE Full text] [Medline: 18693866]
4. Bhavnani V, Fisher B, Winfield M, Seed P. How patients use access to their electronic GP record--a quantitative study. Fam Pract 2011 Apr;28(2):188-194 [FREE Full text] [doi: 10.1093/fampra/cmq092] [Medline: 21084568]
5. Pillemer F, Price RA, Paone S, Martich GD, Albert S, Haidari L, et al. Direct release of test results to patients increases patient engagement and utilization of care. PLoS One 2016;11(6):e0154743 [FREE Full text] [doi: 10.1371/journal.pone.0154743] [Medline: 27337092]
6. O'Kane M, Freedman D, Zikmund-Fisher BJ. Can patients use test results effectively if they have direct access? BMJ 2015 Feb 11;350:h673. [Medline: 25673132]
7. Krist AH, Woolf SH. A vision for patient-centered health information systems. JAMA 2011 Jan 19;305(3):300-301. [doi: 10.1001/jama.2010.2011] [Medline: 21245186]
8. Zikmund-Fisher BJ, Exe NL, Witteman HO. Numeracy and literacy independently predict patients' ability to identify out-of-range test results. J Med Internet Res 2014;16(8):e187 [FREE Full text] [doi: 10.2196/jmir.3241] [Medline: 25135688]
9. Kelman A, Robinson CO, Cochin E, Ahluwalia NJ, Braverman J, Chiauzzi E, et al. Communicating laboratory test results for rheumatoid factor: what do patients and physicians want? Patient Prefer Adherence 2016 Dec;10:2501-2517 [FREE Full text] [doi: 10.2147/PPA.S104396] [Medline: 28008236]
10. Giardina TD, Baldwin J, Nystrom DT, Sittig D, Singh H. Patient perceptions of receiving test results via online portals: a mixed-methods study. J Am Med Inform Assoc 2017:in press (forthcoming). [doi: 10.1093/jamia/ocx140] [Medline: 29240899]
11. Solomon J, Scherer A, Exe N, Witteman H, Fagerlin A, Zikmund-Fisher B. Is this good or bad? Redesigning visual displays of medical test results in patient portals to provide context and meaning. In: CHI'16 Extended Abstracts. 2016 Presented at: Human Factors in Computing Systems 2016; May 7-12, 2016; San Jose, CA p. 2314-2320. [doi: 10.1145/2851581.2892523]
12. Zikmund-Fisher BJ, Scherer AM, Witteman HO, Solomon JB, Exe NL, Tarini BA, et al. Graphics help patients distinguish between urgent and non-urgent deviations in laboratory test results. J Am Med Inform Assoc 2017 May 01;24(3):520-528 [FREE Full text] [doi: 10.1093/jamia/ocw169] [Medline: 28040686]
13. Hsee CK, Loewenstein GF, Blount S, Bazerman MH. Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. Psychol Bull 1999;125(5):576-590. [doi: 10.1037/0033-2909.125.5.576]
14. Hsee CK, Zhang J. General evaluability theory. Perspect Psychol Sci 2010 Jul;5(4):343-355. [doi: 10.1177/1745691610374586] [Medline: 26162182]
15. Stata Corporation. Stata Statistical Software. College Station, TX: Stata Corporation; 2015.
16. Piva E, Plebani M. Interpretative reports and critical values. Clin Chim Acta 2009 Jun;404(1):52-58. [doi: 10.1016/j.cca.2009.03.028] [Medline: 19306860]

Abbreviations

ALT: alanine aminotransferase
ANOVA: analysis of variance
EHR: electronic health record

Edited by G Eysenbach; submitted 06.09.17; peer-reviewed by W Cha, T Risling; comments to author 02.12.17; revised version received 04.01.18; accepted 10.01.18; published 26.03.18.

Please cite as:
Zikmund-Fisher BJ, Scherer AM, Witteman HO, Solomon JB, Exe NL, Fagerlin A
Effect of Harm Anchors in Visual Displays of Test Results on Patient Perceptions of Urgency About Near-Normal Values: Experimental Study
J Med Internet Res 2018;20(3):e98
URL: http://www.jmir.org/2018/3/e98/
doi:10.2196/jmir.8889
PMID:29581088

Original Paper

Novel Interventions and Assessments Using Patient Portals in Adolescent Research: Confidential Survey Study

Lindsay A Thompson[1,2], MS, MD; Rebeccah Mercado[1], MA; Thomas Martinko[1], MD; Ratna Acharya[1], MD

[1]Department of Pediatrics, University of Florida, Gainesville, FL, United States

[2]Department of Health Outcomes and Biomedical Informatics, University of Florida, Gainesville, FL, United States

Corresponding Author:
Lindsay A Thompson, MS, MD
Department of Pediatrics
University of Florida
1699 SW 16th Ave
Gainesville, FL,
United States
Phone: 1 352 627 9323
Email: lathom@ufl.edu

Abstract

Background: While adolescents can receive confidential health care without parental or guardian notification, they are rarely asked about their experiences and opinions regarding their care because participation in research often requires parental consent. Anonymous research with adolescents via confidential patient portals may ameliorate this research gap.

Objective: Because use of a confidential online adolescent patient portal is high at our academic institution, we hypothesized that adolescents would also respond to survey-based research via the portal, especially if asked anonymously and without parental consent. We used a clinical scenario of needing to better understanding adolescent and young adults' views about their health and health care, including information on a long-acting reversible contraceptive (LARC) to test if and how they will use a portal for research.

Methods: Upon receiving Institutional Review Board approval, we sent 2 portal-based surveys about confidential services to 2 groups of females, ages 14 to 25 years, who had attended an adolescent clinic in the past 3 years. This clinic mostly serves Medicaid recipients (80%) and is racially and ethnically diverse with half of patients identifying as African American and roughly 10% Hispanic. The control group was a random sample of female patients who never received a LARC (n=150) and the intervention group included all female adolescents who had received a LARC from the same clinic (n=107). This second sample was manually cross-checked to confirm they had an office visit for this reason. Consenting for themselves, the control group received an email through the patient portal with a link and a request to perform an assessment. The survey for the control group included items assessing health literacy and health communication preferences. The survey for the intervention group included health literacy items as well as items to assess their opinions and perceptions regarding LARCs. We tracked click-through rates and opened messages; each participant received 4 reminders.

Results: While only 3 participants fully completed either survey, email read rates (29/107 [27.1%] of LARC recipients and 39/150 [26.0%] of controls) were encouraging. Additionally, of those who opened the messages, almost twice as many of the LARC recipients (10/107 [9.3%]) read through the entire survey, while less than half read the entire survey as compared to those who received the survey asking about health literacy and health care preferences (6/150 [4.0%]).

Conclusions: The methodology of using adolescent portals for online surveys provides a new avenue for research even though the study did not yield sufficient participation to understand these adolescents' preferences. Future studies need to test if a different survey topic would engage adolescents or if other methods like text-based reminders would improve participation.

(J Med Internet Res 2018;20(3):e101) doi:10.2196/jmir.8340

KEYWORDS

adolescent health services; preventive health services; health information technology

Introduction

Adolescent health research often is overlooked or delayed due to the difficulties of maintaining appropriate confidentiality and privacy [1]. Adolescents can receive confidential health care covering family planning and sexually transmitted disease management without parental or guardian notification, but asking them about their experiences and opinions regarding that same care often requires parental consent. Because this ironic process risks breaking confidentiality, many research questions remain unanswered.

Research advances need to include developing novel and ethical methods of asking questions directly to adolescents. These methods should use caution and follow recommended observational research guidelines such as those created by Ruiz-Canela et al [1] that provide a decision tree to guide researchers, institutional review boards, and ethics committees on how to appropriately enroll adolescents in observational research studies. These authors support confidential research with adolescents without parental consent when 2 conditions are met: when risk is minimal and when there are "specific circumstances that might contribute to vulnerability" [1]. According to the Society for Adolescent Health and Medicine, confidential or anonymous survey research should be considered low risk in the adolescent population as with adults, specifically since deferring parental consent will avoid biasing results [2]. Requiring parental consent on surveys discussing protected information with adolescents potentially strains the parental relationship [2]. Additionally, breaches could break state laws that almost universally allow adolescents to seek family planning interventions without parental approval. For example, a study that required parental consent to ask adolescents about birth control methods would break legal confidentiality. Waivers of parental consent become logical and legal.

Consequently, researchers and adolescent specialists have urged development of alternative methods and locations for adolescent research, particularly survey research, to protect adolescent confidentiality. To date, these have primarily included school- and community-based studies [2]. To our knowledge, online methods of research, specifically confidential patient portals, have yet to be studied. Adolescent portals have become increasingly popular for direct and confidential doctor-patient communication, especially those that maximize confidentiality and aim to educate adolescents on how to access and advocate for their own health needs [3]. Via a private, well-used online patient portal tailored to adolescents' needs [4], we aimed to measure if and how adolescents use their patient portal to consider research participation. By anonymously seeking opinions without parental consent, we sought a research mechanism by which we could solicit adolescents' knowledge and opinions about their health and health care they received, especially if they received a long-acting reversible contraceptive (LARC) whose insertion is protected by state confidentiality and family planning laws.

Methods

Study Population

We received institutional review board approval for this study to recruit young adults and adolescents via a waiver of parental consent. The target population was females aged 14 to 25 years who attended the University of Florida Adolescent Clinic between April 1, 2013, and March 31, 2016. This clinic mostly serves Medicaid recipients (80%) and is diverse, with patients evenly divided between African American and white adolescents and roughly 10% Hispanic adolescents.

Online Patient Portal

In brief, this private adolescent portal was specifically designed for confidential communication between provider and adolescent and does not include parental access unless the adolescent specifically desires it [3,4]. It is widely used in this health care system, with over 60% of adolescents having an activated online portal [3].

Survey Development

We designed 2 similar surveys for implementation that included demographics, the Newest Vital Sign [5], the Single Item Literacy Screener [6], and the Health-Care Self-Determination Theory Questionnaire [7]. For those who had received LARCs in our clinic, we adapted published questions on youth knowledge, experiences, and attitudes about LARC (personal communication with J Peipert, MD, October, 2015) [8,9]. None of the questions required an answer to proceed. We pilot-tested the survey with 16 college students aged 18 to 25 years to ensure the questions were generally understandable. Recommended revisions only changed about 5% of the questions, so we did not perform additional iterations of the evaluation. Both surveys had an 8th grade reading level and would take no longer than 20 minutes to complete.

Identification of Participants

We identified 2 patient populations via the university's online database. First, a random sample of female patients aged 14 to 25 years who had attended the Adolescent Clinic but never received a LARC was identified and generated by the electronic database (n=150). Second, we requested the census of all female adolescents who had received a LARC from the same clinic (n=107). This second sample was manually cross-checked to verify that these individuals had an office visit for this reason to avoid wrongful survey assignment (all had in fact received a LARC insertion).

Survey Implementation

Using the patient portal, clinic physicians sent automated email messages to potential participants. Once logged in to the portal, adolescents received an invitation and a link to the designated survey and consent form that assured anonymity. Given that adolescents may not check email frequently, we sent 4 reminders between August and October 2016. While the survey was anonymous, all respondents were offered a $5 email gift card for survey completion with each email and in the informed consent document, whereby they provided an email address on a different website. For the secondary aim of examining if

adolescents might use portals for survey research, we tracked the rates of click-through messages read and surveys opened to see how many adolescents accessed and opened messages for participation but did not complete the survey.

Results

Table 1 summarizes the similar mean age and usage of possible participants within each group. Only 3 participants fully completed either survey, making it impossible to form inferences about their opinions. However, the process for performing research on adolescents remains possible as evidenced by the number of participants who read at least 1 email message (29/107 [27.1%] of LARC recipients and 39/150 [26.0%] of controls, Table 2). Additionally, of those who opened the messages, almost half of the LARC recipients (10/107, 9.3%) read through the entire survey, while less than half (6/150, 4.0%) did of those who received the survey asking about health and health care preferences.

Table 1. Characteristics of potential survey participants.

Characteristics[a]	Long-acting reversible contraception recipients (n=107)	Age-matched controls without long-acting reversible contraception (n=150)
Female, n (%)	107 (100)	150 (100)
Age, years, mean (range)	19.0 (16-24)	18.6 (14-25)
Younger than 18 years, n (%)	23 (21.5)	53 (35.3)
With self-activated online portal, n (%)	89 (83.1)	107 (71.3)
Age of those with activated online portal, mean	19.0	18.6
Younger than 18 years with self-activated online portal, n (%)	21 (20.0)	51 (33.7)
Age of those without self-activated online portal, mean	18.9	18.8
Younger than 18 years, no self-activated online portal, n (%)	31 (29.4)	60 (40.0)

[a]None of these comparisons is statistically significant.

Table 2. Online portal activity of potential survey participants.

Characteristics[a]	Long-acting reversible contraception recipients (n=107)	Age-matched controls without long-acting reversible contraception (n=150)
Activated online portal (patient-dependent step), n (%)	89 (83.1)	107 (71.3)
Number of messages sent, n	525	600
Number who received up to 4 messages, n (%)	89 (83.1)	104 (97.2)
Number who read at least 1 message, n (%)	29 (27.1)	39 (26.0)
Number who logged into the online portal, n (%)	22 (20.5)	27 (18.0)
Number of days until login after first message sent, n		
0-7 days	7	8
8-15 days	7	4
>16 days	8	15
Number who opened and reviewed survey, n (%)	10 (9.3)	6 (4.0)
Number who completed the survey, n (%)	2 (1.9)	1 (0.7)

[a]None of these comparisons is statistically significant.

Discussion

Principal Findings

This study demonstrates a first step toward using an adolescent patient portal, originally designed for confidentiality, as a mechanism for promoting survey research directly to adolescents. We are encouraged that so many of the adolescents read the messages sent to them and were at least willing to open the patient portal to access the survey even if we cannot form meaningful conclusions about the adolescent and youth points of view. Click-through rates revealed that approximately a quarter of adolescents did in fact read the message, and almost half of those read the survey; they simply did not want to complete the survey provided. It is possible that completion rates may have been higher if we had used a shorter survey or a topic that the adolescents found more engaging. This study confirms that response rates in adolescents are difficult to predict and are likely to be even lower when covering sensitive topics [10]. While we cannot comment on this population's health literacy or opinions on LARCs, we can confirm that adolescents responded to email solicitation for patient portal participation, leaving opportunities open for future research. We believe such research will finally bridge the gap between needing to

understand adolescent opinions and maintaining appropriate confidentiality.

Limitations

There are several limitations that warrant discussion. First, the number of adolescents who have active portal accounts and opened messages in this sample may be higher than the general population due to constant promotion of the adolescent portal in this clinical setting [4]. Future studies would need to take portal activation rates into consideration. Conversely, given the high activation rate in this population, these adolescents may have had concerns that their provider would learn information about them that they did not want them to have. Second, having a larger and more diverse sample may yield meaningful completion rates; we were limited by the number of female youths who had received LARCs. Third, the topic of the survey may not have been engaging enough for the adolescents, and a different survey topic may achieve greater response and completion. Finally, the email solicitation or series of clicks that adolescents had to perform to reach the survey may have been inhibitive. Some electronic health records have internal survey-building capacity or text-based options and should be encouraged as future avenues for research.

Conclusion

Online portals offer an important potential as a medium for adolescent research, but topic selection and methods of engagement need to be refined. The electronic health record system at this health institution will soon begin offering texts from the portal and internal surveys, potentially increasing adolescent response rates. Future research should ask adolescents, through interviews or focus groups, especially those who were identified for participation for this study, what mechanisms and content they prefer when discussing confidential topics and what possible barriers and facilitators they perceive.

Acknowledgments

We would like to express our deepest gratitude to the following individuals within our organizations for their substantial contributions to this project: the University of Florida team including Gloria Pflugfelder Lipori, who was integral in the development and execution of the surveys, and Nazokat Usmanova, who contributed to the development of the surveys and processes of using the portal for survey research. We received no funding to support this project.

Conflicts of Interest

None declared.

References

1. Ruiz-Canela M, Lopez-del Burgo C, Carlos S, Calatrava M, Beltramo C, Osorio A, et al. Observational research with adolescents: a framework for the management of the parental permission. BMC Med Ethics 2013 Jan 03;14:2 [FREE Full text] [doi: 10.1186/1472-6939-14-2] [Medline: 23286743]
2. Santelli JS, Smith RA, Rosenfeld WD, DuRant RH, Dubler N, Morreale M, Society for Adolescent Medicine. Guidelines for adolescent health research. A position paper of the Society for Adolescent Medicine. J Adolesc Health 2003 Nov;33(5):396-409. [Medline: 14596961]
3. Thompson LA, Novak M, Schentrup AM. Adolescent health and the electronic health record: can it be a social media tool for quality adolescent care? Adolesc Med State Art Rev 2014 Dec;25(3):698-710. [Medline: 27120894]
4. Thompson LA, Martinko T, Budd P, Mercado R, Schentrup AM. Meaningful use of a confidential adolescent patient portal. J Adolesc Health 2016 Feb;58(2):134-140. [doi: 10.1016/j.jadohealth.2015.10.015] [Medline: 26802988]
5. Warsh J, Chari R, Badaczewski A, Hossain J, Sharif I. Can the Newest Vital Sign be used to assess health literacy in children and adolescents? Clin Pediatr (Phila) 2014 Feb;53(2):141-144. [doi: 10.1177/0009922813504025] [Medline: 24065737]
6. Morris NS, MacLean CD, Chew LD, Littenberg B. The Single Item Literacy Screener: evaluation of a brief instrument to identify limited reading ability. BMC Fam Pract 2006 Mar 24;7:21 [FREE Full text] [doi: 10.1186/1471-2296-7-21] [Medline: 16563164]
7. Deci EL, Ryan RM. Intrinsic Motivation and Self-Determination in Human Behavior. New York: Plenum; 1985.
8. Peipert JF, Zhao Q, Allsworth JE, Petrosky E, Madden T, Eisenberg D, et al. Continuation and satisfaction of reversible contraception. Obstet Gynecol 2011 May;117(5):1105-1113 [FREE Full text] [doi: 10.1097/AOG.0b013e31821188ad] [Medline: 21508749]
9. Abraham M, Zhao Q, Peipert JF. Young age, nulliparity, and continuation of long-acting reversible contraceptive methods. Obstet Gynecol 2015 Oct;126(4):823-829. [doi: 10.1097/AOG.0000000000001036] [Medline: 26348177]
10. Wyrick DL, Bond L. Reducing sensitive survey response bias in research on adolescents: a comparison of web-based and paper-and-pencil administration. Am J Health Promot 2011;25(5):349-352. [doi: 10.4278/ajhp.080611-ARB-90] [Medline: 21534838]

Abbreviations

LARC: long-acting reversible contraceptive

Edited by G Eysenbach; submitted 03.07.17; peer-reviewed by G Strudwick, M Habibi-koolaee, G Zieve, S Whitehouse; comments to author 23.11.17; revised version received 16.01.18; accepted 27.01.18; published 21.03.18.

Please cite as:
Thompson LA, Mercado R, Martinko T, Acharya R
Novel Interventions and Assessments Using Patient Portals in Adolescent Research: Confidential Survey Study
J Med Internet Res 2018;20(3):e101
URL: http://www.jmir.org/2018/3/e101/
doi:10.2196/jmir.8340
PMID:29563077

Original Paper

Developing a Shared Patient-Centered, Web-Based Medication Platform for Type 2 Diabetes Patients and Their Health Care Providers: Qualitative Study on User Requirements

Gerda Bernhard[1], MSc; Cornelia Mahler[1], Dr sc hum, MA, RN; Hanna Marita Seidling[2,3], PD, Dr sc hum; Marion Stützle[2,3], Dr rer nat; Dominik Ose[1,4], PD, DrPH; Ines Baudendistel[1], Dr sc hum, MSc; Michel Wensing[1], PhD, MSc; Joachim Szecsenyi[1], Dr med, MD, MSc

[1]Department of General Practice and Health Services Research, Heidelberg University Hospital, Heidelberg, Germany

[2]Cooperation Unit Clinical Pharmacy, Heidelberg University Hospital, Heidelberg, Germany

[3]Department of Clinical Pharmacology and Pharmacoepidemiology, Heidelberg University Hospital, Heidelberg, Germany

[4]Division of Cancer Population Sciences, Department of Population Health Sciences, University of Utah, Salt Lake City, UT, United States

Corresponding Author:
Gerda Bernhard, MSc
Department of General Practice and Health Services Research
Heidelberg University Hospital
Marsilius-Arkaden, Turm West, Im Neuenheimer Feld 130.3
Heidelberg, 69120
Germany
Phone: 49 (0) 6221 56 3438
Fax: 49 (0) 6221 56 1972
Email: gerda.bernhard@med.uni-heidelberg.de

Abstract

Background: Information technology tools such as shared patient-centered, Web-based medication platforms hold promise to support safe medication use by strengthening patient participation, enhancing patients' knowledge, helping patients to improve self-management of their medications, and improving communication on medications among patients and health care professionals (HCPs). However, the uptake of such platforms remains a challenge also due to inadequate user involvement in the development process. Employing a user-centered design (UCD) approach is therefore critical to ensure that user' adoption is optimal.

Objective: The purpose of this study was to identify what patients with type 2 diabetes mellitus (T2DM) and their HCPs regard necessary requirements in terms of functionalities and usability of a shared patient-centered, Web-based medication platform for patients with T2DM.

Methods: This qualitative study included focus groups with purposeful samples of patients with T2DM (n=25), general practitioners (n=13), and health care assistants (n=10) recruited from regional health care settings in southwestern Germany. In total, 8 semistructured focus groups were conducted. Sessions were audio- and video-recorded, transcribed verbatim, and subjected to a computer-aided qualitative content analysis.

Results: Appropriate security and access methods, supported data entry, printing, and sending information electronically, and tracking medication history were perceived as the essential functionalities. Although patients wanted automatic interaction checks and safety alerts, HCPs on the contrary were concerned that unspecific alerts confuse patients and lead to nonadherence. Furthermore, HCPs were opposed to patients' ability to withhold or restrict access to information in the platform. To optimize usability, there was consensus among participants to display information in a structured, chronological format, to provide information in lay language, to use visual aids and customize information content, and align the platform to users' workflow.

Conclusions: By employing a UCD, this study provides insight into the desired functionalities and usability of patients and HCPs regarding a shared patient-centered, Web-based medication platform, thus increasing the likelihood to achieve a functional and useful system. Substantial and ongoing engagement by all intended user groups is necessary to reconcile differences in requirements of patients and HCPs, especially regarding medication safety alerts and access control. Moreover, effective training of patients and HCPs on medication self-management (support) and optimal use of the tool will be a prerequisite to unfold the platform's full potential.

(J Med Internet Res 2018;20(3):e105) doi:10.2196/jmir.8666

KEYWORDS

medication; self-management; patient portal; patient participation; type 2 diabetes mellitus; focus groups; primary care

Introduction

Medication Self-Management in Type 2 Diabetes

The prevalence of type 2 diabetes mellitus (T2DM) is steadily growing and represents a challenge for health care systems worldwide [1]. The complex nature of T2DM management requires ongoing efforts by the patient in collaboration with health care professionals (HCPs) and informal support networks. Compared with other chronic conditions, T2DM management requires extensive self-management behaviors of patients, such as adhering to often complex medication regimens, lifestyle modifications (eg, diet, physical activity), recognizing and responding to symptoms, and managing acute episodes [2]. Patient education and promoting patients' self-management has long been recognized as an important strategy in diabetes management [3].

Medication self-management, defined as the various tasks patients must undertake to effectively manage their therapeutic regimen and sustain safe medication use over the long term [4], is a critical skill for patients with T2DM. Considering this definition, taking medications is a complex, multistep task and requires patients filling the prescription, understanding the medication regimen, organizing and correctly taking the medication, monitoring the intake, and then sustaining medication use over time [4]. Research has demonstrated that patients with T2DM often have an inadequate understanding on how to safely take their medications and concerns regarding the appropriateness and safety of their regimen [5,6]. Inadequate understanding can lead to improper use, adverse drug events (ADEs), and suboptimal adherence [7-9]. In fact, suboptimal medication adherence is common among patients with T2DM [10] and is associated with poor health outcomes [11].

Health Information Technology

Health information technology (HIT) that is patient-centered [12], such as Web-based personal health records, creates new opportunities to facilitate patients' diabetes and medication self-management and enhances patient outcomes [13-15]. The use of these technologies can provide patients access to personal medication information and essential resources to facilitate informed decision making, promote communication between patients and HCPs, and enhance patient engagement and self-management [14,16], particularly in patients with chronic conditions such as T2DM [16,17]. However, despite the potential benefits of HIT, introducing new technology to health care has proven difficult [18] and adoption rates are often low [19]. Factors identified in the literature that inhibit successful HIT implementation include inadequate funding, lack of IT infrastructure, poor leadership, unrealistic timeline, and inadequate end-user engagement and input [18,20]. Hence, research emphasized the need to directly involve intended users in requirement specification, design, and testing to ensure that HIT matches users' cognitive abilities and needs, and support self-managed care [19].

User Requirements Elicitation

Utilizing a user-centered design (UCD) [21] will enhance closer user participation throughout the entire development process and result in better tailoring to user requirements and needs [22,23]. Likewise, the FITT framework (Fit between Individuals, Task, and Technology) highlighted that successful adoption of a technology depends on the fit between the attributes of the individual user, the task, and the technology [24].

Today, a vast number of mobile apps exist to help manage outpatient medication use for diverse medical conditions. A systematic review [25] found that most apps provided medication reminders and half enabled creating a medication history, list, or log, whereas only few helped patients to organize their regimen and check for drug interactions. Overall, the quality, content, and functionality varied greatly. Common user criticism revealed technical malfunctions, poor compatibility with complex or varying regimes, and absence of desired features. The review concluded that further research is necessary to improve the design, content, and features from a patient perspective [25]. To date, only a few studies [14,26,27] have reported user requirements regarding the design and development of patient-centered HIT to support T2DM patients in self-managing their medications [28]. Requirements reported in these studies included reliable information on medication side effects and interactions, electronic messaging, selectively disclosing information, refill reminder functionality, and a user-friendly format [14,26,27]. Although these studies provide valuable information on user requirements, they have not elicited the views of both user groups: patients and HCPs. In Germany, diabetes care is largely provided in primary care with general practitioners (GPs) and health care assistants (HCAs) playing a key role in coordinating care, pharmaceutical treatment, and patient education. Thus, involving patients as well as GPs and HCAs as collaborative partners will help to identify what functionalities users require to accomplish the medication management tasks. Moreover, previous research has highlighted that HCPs endorsement of HIT is pivotal to increase patients' participation and sustain use over time [29,30]. E-medication is a primary objective in Germany's eHealth strategy; however, the development of a national e-medication infrastructure is slow and cumbersome [31].

Therefore, this study employed a UCD to identify necessary requirements in terms of functionalities and usability of a shared patient-centered, Web-based medication platform from the perspective of patients with T2DM and their HCPs (GPs and HCAs). Patients should be able to access, share, manage, and maintain personal medication-related information through this platform, with the intention of enhancing patients' knowledge and strengthening patients' active participation, thus helping

them to better self-manage their medications and support safe medication use.

Methods

Research Design

A qualitative study based on focus groups was employed to explore the views of patients with T2DM and HCPs on their requirements of a shared patient-centered, Web-based medication platform [32,33]. Focus groups provide the opportunity for a group of people to explore and clarify their perspectives than would arise in individual interviews, facilitated by social interaction. This qualitative method is thus particularly useful for exploring not only what people think but also how they think and why they think that way [32]. The study was conducted as part of a larger German research project called INFOPAT (Information Technologies for Patient-Centered Health Care, 2012-2016) [34] studying the needs of chronically ill patients and their HCPs to develop tailored information technologies (ie, Web-based medication platform) and a medication communication intervention to facilitate medication self-management, improve medication safety, and patient-provider communication. Recruitment and study procedures have also been described in detail previously [35,36].

Study Participants and Recruitment

Between April and July 2013, participants were purposefully recruited from the Rhine-Neckar region in southwestern Germany. The purposive sampling strategy aimed to achieve variation in patient characteristics, including education, duration of diabetes, and medication regimen to elicit a broad range of experiences and requirements. German or Turkish-speaking adults (largest ethnic minority in Germany) with a T2DM diagnosis who were self-administering prescribed diabetes medications (oral hypoglycemic agents [OHA] only or insulin only or OHA and insulin) were approached personally through 3 channels: local self-help groups, GP practices, and during routine appointments at the Heidelberg University Hospital. There were no restrictions on age, nor was computer or internet experience a prerequisite to take part in the study.

GPs and HCAs with diabetes expertise, and experienced in caring for T2DM patients, were recruited by a letter through a list of cooperating academic teaching and research practices from different geographic locations. The sampling approach aimed to ensure diversity in terms of practice size, urban/rural location, and computerization in practice. All participants gave informed written consent before study enrollment. Ethical approval was granted by the Ethics Committee of the Medical Faculty of Heidelberg University (no S-673/2012). Participants received a compensation of 50 €.

Focus Groups

From May until July 2013, 8 focus groups with a total of 48 individuals, including 25 patients with T2DM, 13 GPs, and 10 HCAs, were conducted. Focus groups were conducted with 6-8 participants per group. Of these 8 groups, 4 groups included only T2DM patients, 3 groups included GPs and HCAs, and 1 group included only GPs. Each session was facilitated in German by an experienced moderator and comoderator (authors GB, CM, or DO) with the assistance of a trained note taker. The meeting with the group of patients of Turkish descent was simultaneously translated into Turkish (due to limited German language proficiency) by a bilingual Turkish project partner who acted as a comoderator. Participants completed a brief sociodemographic questionnaire anonymously in conjunction with the focus group. Semistructured, pilot-tested interview guides alongside a moderator guide were used to guide the discussion. Interview guides were matched on key themes and covered participants' requirements and needs regarding a shared patient-centered, Web-based medication platform. Topics covered in the focus groups, as they pertain to this paper, included open-ended questions and probes to encourage a broad discussion about participants' experiences with their medication management, attitudes, and opinions toward using a Web-based shared medication platform and its technological and content requirements (see Multimedia Appendix 1 for sample questions). At the beginning of each focus group, the moderator briefly presented the general idea of a medication platform by a PowerPoint presentation. Postfocus group debriefings were conducted, and central themes were documented in a research diary. Each focus group lasted between 110 and 130 min, and was audio- and videotaped. Recruitment of new participants ceased when no new themes emerged in the group discussions [37].

Characteristics of Participants

Most patients had complex medication regimens (≥5 different types of medications taken regularly per day, not restricted to diabetes medication) and had on average 3 other chronic conditions. Of 25 patients, 16 (64%) reported having access to a personal computer/laptop and about half stated to use the internet for medication-related information. Almost all HCPs had an internet connection in their practice, and 8 out of 13 GPs (62%) reported to use electronic decision support systems. Tables 1 and 2 present demographic characteristics, computer/internet use, and recruitment of patients with T2DM and HCPs.

Data Analysis

Audio- and video recordings were fully transcribed by trained staff with anonymity of participants completely protected and reviewed by the moderator (GB) for accuracy. The observers' notes and debriefing notes were synthesized and integrated into the data analysis process. Data were analyzed iteratively using qualitative content analysis to structure material in codes (labels of condensed meaning units), subcategories, and categories (themes) [38-40].

Table 1. Characteristics of patients with type 2 diabetes who participated in the focus groups.

Patient characteristics	Patient focus groups (N=25)
Gender (female), n (%)	7 (28)
Age (years), mean (SD); range	64 (8.6); 49-77
Diabetes duration (years), mean (SD); range	13.9 (10.6); 0.8-38
Number of other chronic conditions, mean (SD); range	3.4 (1.6); 1-7
First language, n (%)	
German	18 (72)
Turkish	7 (28)
Number of different medications taken regularly per day[a] , n (%)	
1-2 medications	2 (8)
3-4 medications	6 (24)
5-6 medications	5 (20)
≥7 medications	12 (48)
Diabetes medication, n (%)	
Oral hypoglycemic agents only	13 (52)
Insulin only	3 (12)
Oral hypoglycemic agents and insulin	9 (36)
Education/school years, n (%)	
Secondary school (9 years)	12 (48)
Secondary modern school (10 years)	5 (20)
Grammar school (13 years)	8 (32)
Computer/laptop at home, n (%)	16 (64)
Internet use at home, n (%)	14 (52)
Internet use for medication-related information, n (%)	12 (48)
Recruitment through, n (%)	
Self-help groups	15 (60)
Heidelberg University Hospital	6 (24)
General practitioner practices	4 (16)

[a]Not restricted to diabetes medication.

Development of thematic categories was guided by priori objectives identified in the interview guide while also allowing new themes to emerge from the data [40]. Moreover, 2 researchers (GB, CM) independently read transcripts and notes thoroughly and then coded data to establish subcategories and categories through consensus. At first, transcripts were deductively analyzed by assigning initial categories corresponding to the interview guide. Next, material pertaining to each category was analyzed inductively to refine subcategories. If the data revealed new information not fitting the preliminary coding scheme, categories were developed inductively. Throughout the iterative process of revisiting the data and connecting them with new insights, an initial coding scheme was established [40].

The researchers used Atlas.ti (Version 7.0.80, Scientific Software Development GmbH, Berlin, Germany), a qualitative software package, for organizing and coding the data. Researchers met regularly throughout the study to discuss categories and subcategories until consensus on the final set of categories was reached. By taking into account the number of focus groups reporting specific requirements, prioritization of requirements was possible.

Table 2. Characteristics of participating health care professionals. DMP: disease management program. N/A: not applicable. PC: personal computer.

Health care professional characteristics	Professional focus groups	
	General practitioners (N=13)	Health care assistants (N=10)
Gender, (female), n (%)	6 (46.2)	10 (100)
Age (years), mean (SD); range	54.1 (9.2); 35-64	38.6 (11.8); 21-52
Structure of practice, n (%)		
Solo practice	4 (30.8)	4 (40)
Group practice	7 (53.8)	6 (60)
Practice sharing	1 (7.7)	
Ambulatory health center	1 (7.7)	
Location of practice, n (%)		
City center	6 (46.2)	5 (50)
Suburbia	5 (38.5)	1 (10)
Rural area	2 (15.4)	4 (40)
Years of work experience, mean (SD); range	24.5 (9.8); 6-40	15.5 (12.5); 0-35
Participation in DMP diabetes type 2, n (%)	13 (100)	10 (100)
Solely electronic documentation, n (%)	6 (46.2)	3 (30)
Use of electronic decision support systems, n (%)	8 (61.5)	N/A
Internet connection in practice, n (%)	12 (92.3)	9 (90)
PC with practice software connected to internet, n (%)	9 (69.2)	8 (80)
Recruitment through, n (%)		
Academic teaching practices	12 (92.3)	8 (80)
Research practices	1 (7.7)	2 (20)

Results

User Requirements Regarding Functionalities and Usability of a Shared Patient-Centered, Web-Based Medication Platform

Focus group participants discussed their requirements in terms of functionalities and usability of a shared patient-centered, Web-based medication platform. Categories and illustrative quotes are presented in more detail in the following section. Overall, GPs and HCAs had similar requirements regarding the medication platform as the great majority of codes were mentioned by both groups. Thus, data from GPs and HCAs were pooled together. For publication, the coding scheme and quotations were translated into English by the first author (German-native and fluent speaker of English) and thereafter cross-checked by an English- and German-native speaking coauthor (CM). Unique identifiers are used to protect participants' anonymity (P, patient; GP, general practitioner; HCA, health care assistant; FG, focus group). To facilitate readability, categories, subcategories, and associated codes are presented in Tables 3 and 4. Moreover, requirements were prioritized (+-++++) based on the number "(1-4)" of patient and HCP focus groups reporting a specific requirement.

Functionalities of the Medication Platform

Participants' expectations regarding functionalities of the medication platform were divided into 5 subcategories: (1) security, access control, and supported data entry; (2) safety alerts, reminders, and notifications; (3) tracking medication history; (4) support features; and (5) electronic messaging and information sharing (see Table 3 and Multimedia Appendix 2). Although patients and HCPs had mostly similar expectations regarding functionalities, they had controversial views on automatic interaction checks and safety alerts for patients and on patients' ability to control access to the platform.

Table 3. Required functionalities of the medication platform.

Subcategory and code	Patient[a]	HCP[b]
Security, access control, and supported data entry		
Data security and privacy	++++	++++
Rapid access in case of emergency	++	++
Patient can customize and restrict access to platform	+++[c]	
Physicians and HCPs need full access		+++[c]
Restrict entering and changing information in platform		+++
Simple data upload, automatic spell, and plausibility check	+++	+++
Interoperability with management software systems		++++
Safety alerts, reminders, and notifications		
Automatic interaction checks and safety alerts, trigger alert messages and visual clues to highlight interactions, risks and contraindications, what to do and specific instructions for safe use	++++[c]	
Only high-severity drug-drug interactions, allergy alerts, contraindications, duplicate medications, and what to do (HCP perspective: physician judgment is needed)		++++[c]
Highlight potentially hazardous medications and provide specific precautions	++	+
Signalize new entries and changes made (eg, pop-up, colored)	++	++
E-reminder to undertake medication reconciliation, counseling, and review of therapy		++
E-reminders to support medication intake or discontinue intake	+	
Tracking medication history		
Complete medication regimen	+++	++++
Date of prescription, medication change, and update	++	++
Person who entered or changed information		++
Reason for changes or discontinuing medication	+++	++++
Occurrence of adverse drug events	+++	++++
Medication dispensing information from pharmacy		++
Patients can add specific information (eg, over-the-counter medications, symptoms)		++
Support features		
Search function	++	++
Medication possession calculator	++	+
Insulin dose calculator	+	+
Medication plan and information can be printed, and send electronically	+++	+++
Electronic messaging and information sharing		
Exchange of experiences and information between patients	+	
Electronic messaging between HCPs		+

[a]Requirements of patients with type 2 diabetes, prioritized according to the number (1-4) of focus groups reporting requirement.

[b]Requirements of health care professionals (HCP; general practitioners and health care assistants), prioritized according to the number (1-4) of focus groups reporting requirement.

[c]Controversial views between patients and health care professionals.

Table 4. Requirements regarding usability of the medication platform.

Subcategory and code	Patient[a]	HCP[b]
User interface		
Structured information according to diagnosis or therapeutic indication, long-term and on-demand medication	+++	++++
Structured information in a chronological order	++++	++++
Intuitive design and navigation, tailored to users' workflow	++++	++++
Ergonomic presentation, large font size, customizable adaptation of information density	+++	+++
User-centered provision of information		
Lay and multilingual language, for example, evidence-based information	+++	+++
Glossary to support comprehensibility of medical terms, wiki to answer important questions	++	++
Use of visual aids, clues, and videos to facilitate understanding of information	+++	++
User guide, provision of training, and links to additional support	++	

[a]Requirements of patients with type 2 diabetes, prioritized according to the number (1-4) of focus groups reporting requirement.

[b]Requirements of health care professionals (HCP), prioritized according to the number (1-4) of focus groups reporting requirement.

Security, Access Control, and Supported Data Entry

Data security and privacy issues were intensively discussed across all focus groups, and appropriate security and access methods (eg, secure authentication) were fundamental for patients and HCPs to use the platform. Participants stressed the platform would contain sensitive information on diagnoses and medications, which was potentially valuable for third parties (eg, insurance companies, pharmaceutical industry). Hence, several patients and HCPs mentioned concerns to become "gläsern" (transparent, P2-FG3) and monitored by the platform:

> *I would not use it if it would be cumbersome to use. That would be the first requirement—the handling. And I would not use it if it had negative effects for the patients or if I would expect more control of my work, when I feel that it is going to be a surveillance tool for my work or when I witness that insurances, border authorities or someone else is interested in these data.* [GP2-FG4]

Accordingly, participants stated they had to decide whether the benefits of using the platform would outweigh the theoretical risks. On the other hand, participants requested an emergency access functionality in the platform to enable physicians' rapid access to a patients' current medication list and important patient-related information (eg, allergies, intolerances, risk factors). To be acceptable to them, most patients emphasized they need to be in control of their data and the authorization of different HCPs and significant others to access and add information to their personal account in the platform:

> *The access would have to be very restricted and controlled by myself.* [P2-FG1]

Although most participants supported the general idea of a patient-controlled platform, most HCPs did not support patients' ability to withhold information or restrict access to certain information. HCPs stressed that physicians need to be fully informed about a patient's regimen to make informed decisions. Furthermore, participants had extensive discussions about who should be able to enter and change information in the platform.

Most HCPs thought it was important to restrict entering and changing information in the platform. For instance, physicians can enter their own prescription but should not be able to change medications prescribed and entered by other HCPs.

Furthermore, HCPs suggested that patients can enter over-the-counter medications (OTCs) and symptoms experienced in the platform but should not be allowed to change or delete a physician's prescribed medication. Otherwise, HCPs stressed "I can no longer trust my own case" [GP1-FG4] and perceived this would affect their liability as well as the reliability of information in the platform. Likewise, one HCA emphasized:

> *The patient should have the possibility to add something [to the platform], but it should be clear that it comes from the patient, yes, so one always knows, he [the patient] has added something. But he [the patient] should not be able to delete anything…* [HCA1-FG2]

Besides, HCPs mentioned concerns about the accuracy of patient-entered data. Although some patients stated they wanted to enter and update information themselves, patients with limited computer experience or skills said they wanted their GP, functioning as a coordinator, to enter or upload prescription information. Others stated they would ask relatives or friends to support them entering information and using the platform.

A fundamental requirement for patients and HCPs was that the platform facilitated easy access, entry, and upload of medication information while maintaining high security standards. For instance, both groups perceived an automated entry of medication information was important for ease of use and a prerequisite for acceptance of the platform.

The HCP groups underlined that patients with T2DM often get prescriptions from different providers, for instance, GP, specialist, or hospital, and buy OTCs directly at the pharmacy. Thus, the platform should interoperate with different

management software systems and enable an automatic data upload. Many HCPs stressed they would not enter medication information twice, in their own system and in the medication platform:

> *The upload from the practice management software should be automatic, that it isn't more work to open the platform in the practice management software. I think this is very important because when there is an administrative effort, it becomes difficult, but if there are interfaces to the practice management software, then it's certainly a good thing…* [GP1-FG4]

Some patients suggested adding new medications to the platform could be simplified by scanning the medication barcode with their mobile device or a barcode scanner. To illustrate this suggestion, 1 patient said:

> *We have also heard here that we are sometimes simply overwhelmed with our medications… because they either constantly change or something else. Thus, the system should be designed in a way that it is relatively simple for me to put in my medications. And I have to keep it [medication] up to date. I mean if this is too complicated or takes too much effort… it is not up to date. Then it's of no use… as I said, if I handle the system self-responsible, it must be easy, a medication package has a barcode. Quickly, with the mobile phone the number is scanned. Quickly, the medication is entered. If I have to enter everything each time by hand… that's just too much.* [P5- FG4]

Others emphasized if they entered a medication manually, the platform should feature an automatic spell and plausibility checker and provide an automatic word completion for a quick entry.

Safety Alerts, Reminders, and Notifications

Safety alerts, reminders, and automated notifications were well-discussed in all focus groups with differences detected in patients' and HCPs' expectations. Most patients perceived they lacked information on potential side effects, long-term effects, and drug interactions of their prescribed medication and voiced concerns regarding the safety of their regimen. Thus, many patients reported to seek risk-related information from a range of sources including their GP, specialist, community pharmacist, local self-help group, friends, as well as Web-based (internet) resources.

Accordingly, patients suggested the platform should provide comprehensive risk-related information and automatically check and highlight interactions (eg, drug-drug, allergy, food) in their regimen.

> *I have to take lots of different medications…there [patient information leaflet] you can read about incompatibilities with this and that substance, but I don't even know at all which substance is in which medication…it would be important, if I have my medications in such a system that it automatically reconciles: "Do these fit together at all?" I mean, I always have to trust my doctor that he knows this, but sometimes I have the feeling: "How does he know all of this?" Because there is so much stuff he has to know. Sometimes I have a bad feeling, whether he really knows that…I have an insecurity with the medications…I often have the feeling it is a calculated risk…* [P5-FG4]

Thus, several patients wished real-time safety alerts to pop-up automatically indicating the severity of interactions in their regimen, for instance, by using distinctive color-coding (eg, according to traffic light, red=serious). On the other hand, 2 patients were also averse to receiving information on potential adverse effects, as this may negatively influence their attitude toward a medication.

Likewise, HCPs expressed hesitation about offering an automatic interaction check to patients and mentioned concerns about how patients dealt with this critical information, as in their view, this could increase patient fear, encouraging nonadherence, and numerous discussions. HCPs also noted that interactions in the regimen of chronically ill patients are common, often not avoidable, and needed to be judged by them. Over half of the GPs reported to use electronic decision support systems for prescription writing. Although GPs generally valued drug interaction and allergy alerts during prescribing, they stated to override these frequently due to little clinical significance and extensive numbers of warnings. Nonetheless, HCPs believed that patients need to be informed about the most relevant adverse and long-term effects to monitor their own treatment and know how to reduce potential risks. Thus, HCPs suggested only showing relevant adverse effects, contraindications, duplicate medications, allergy alerts, and high severity drug-drug interactions to patients. At the same time, HCPs, however, also recognized difficulties regarding liability and legal implications when only certain warnings would be displayed in the platform. The following exchange exemplifies this:

> *There are warnings; you have to go deaf, this doesn't help…only the absolute relevant warnings should pop-up…* [GP1-FG1]

Response of another participant:

> *This is not possible, because it is actuarially all relevant, that's the problem.* [GP3-FG1]
>
> *When with every ACE inhibitor and potassium-sparing diuretic it pops-up every time… we know that, and this doesn't help. So, it should really only be what is relevant… so that the absolute no-go's pop-up.* [GP1-FG1]

Both groups pointed out that it was necessary to concisely describe the actions to be taken by patients in lay language in the alert to mitigate potential adverse effects and to promptly contact their treating physician.

A few participants also requested that the platform notified users about potentially hazardous medications (eg, anticoagulation) or medications with unclear benefits. In addition, patients and HCPs thought it was necessary to highlight new entries or changes made in the platform automatically, for instance, by using visual clues (eg, pop-up, color):

> *That you do not have to check constantly if something has been changed, at what time something has been*

changed, but that this actually runs automatically… [HCA3-FG2]

In general, HCPs saw great value of the platform for medication reconciliation and counseling by combining all relevant medication information of a patient in one place. Hence, HCPs were interested in setting up an automated prompt that reminded them periodically to undertake medication reconciliation and counseling. Similarly, some patients wished the platform enabled setting up tailored audible medication-taking prompts that reminded them to take their medications at specific time intervals or, for instance, to discontinue intake before surgery. To enable this feature, patients suggested the platform should be linked to their mobile device.

Tracking Medication History

Tracking a patient's medication history received great attention in the group discussions and was perceived by all participants, especially HCPs, as an essential component of the platform. Participants were enthusiastic about the platforms' potential to increase the ease of documentation and produce a structured presentation of a patient's complete medication history (eg, including prescription and nonprescription medication, supplements), thus facilitating information exchange and medication reconciliation across health care sectors and professionals. HCPs underscored current challenges (eg, resource and time constraints, lack of cross-sectorial collaboration, and technical interoperability, drug discount contracts) in reconciling medications of their chronically ill patients, particularly during transitions of care. For the patient groups, capturing their medication history was relevant to see how their condition and treatment developed over time. Several patients described how they kept track of their currently and previously used medications by creating paper-based medication lists where they documented changes (eg, regarding dosage, frequency, and generic substitution) made in their regimen and related these to clinical parameters (eg, glycated hemoglobin):

What is important for me, I always make a note when I changed the dosage or the medication […] that I can enter it [in the platform] and that I know the dosage has been […] that's a thing I like to check. Has it improved or worsened since the change. I found this was very important and have written it down next to it. [P3-FG1]

Moreover, patients stressed that medication changes are common, often tied to specific problems or therapeutic goals, and without documentation, they would have difficulties keeping track of this information. HCPs underlined often not being fully informed about a patient's medication regimen as only the patients are in the position to account for their self-medication. Thus, patients should enter OTCs and symptoms experienced to complete medication history taking. HCPs, however, experienced that patients had difficulties disclosing their co-usage of OTCs, vitamins, or herbal supplements. Hence, HCPs suggested patients would need prior guidance on the importance of accurately documenting prescription and nonprescription medications in the platform. Above all, participants across focus groups emphasized it was crucial to consistently record reasons for changing or discontinuing medications in a patient's medication history, including information on side effects, ADEs, and intolerances. This information was perceived vitally important to make informed treatment decisions and ensure patient safety. In addition, participants stressed the platform should automatically capture the name of the person (eg, patient, prescribing physician, dispensing pharmacy) entering data as well as the date of each transaction:

I would like to have a medication history, who prescribed and discontinued what, when and why…because often with chronically ill patients there are let's say circular procedures: Medication A, Medication B, Medication C, Medication A and then the question arises: did he not tolerate it? After some time, the patient doesn't know it anymore and I have to admit sometimes I'm not either…to have a comment field to record the reason for discontinuing the medication, due to intolerance, allergy, medication change, hospital stay, ineffectiveness or so on. [GP2-FG4]

Documentation of possible adverse drug reactions with date of occurrence. That there is somewhere a note, there have been adverse effects. After five years I forgot that one [a patient] on Amlodipin got edemas and I will prescribe it again, then it happens again. There are also more severe adverse effects… [GP2-FG2]

If a medication is changed that the date is recorded when it was changed and what was prescribed instead… [HCA2-FG2]

Furthermore, in participants' view, medication dispensing information from the pharmacy (eg, date of dispensing, generic substitution, and initials of dispensing pharmacy) would shed light on generic substitution and may improve patients' comprehension. Thus, the platform should clearly link the patients' prescribed medication with the dispensed generic medication.

Support Features

Participants suggested incorporating support features to facilitate retrieval of information in the platform and to help organize daily medication taking. Due to large amounts of data stored in the platform, participants stressed the platform should contain a search function allowing rapid retrieval of information. To support patients' self-management, many participants wished to be able to print relevant information (eg, medication list, specific directions for use, and administration) as well as to send information electronically. Participants also highlighted they would benefit if the platform offered a feature to calculate medication possession (ie, number of refills remaining) and the amount of insulin units needed to reach a target blood glucose concentration, as the following discussion illustrates:

It would be practical…if the platform would list, when one has to get a refill, so one can plan ahead… [HCA1-FG3]

I have another idea. I have insulin-dependent diabetes and fly to New York…or further [across several time

> *zones] How do I have to adjust my insulin dosage now? There are great calculations models available…* [GP3-FG3]

> *So, I would even say that one could also specify the time period, so that one knows how much he has to take with him…* [HCA1-FG3]

Electronic Messaging and Information Sharing

The potential of the platform to enable electronic messaging among users was only discussed briefly and revealed controversial aspects among focus group participants. A small number of patients said they would like to share information and personal experiences, for example, with certain medications, adverse effects, or alternative treatment options, with their peers through an integrated anonymous chat or online user forum:

> *I have recognized it right from the beginning how each of us told about his illnesses and so on. I heard about this and that. Aha! That's the same with me. There [relating to the platform] it is possible to exchange certain experiences. Surely, I cannot say, […] you have to do this and that. That doesn't work. But you can get suggestions and maybe you see: others are in the same situation. He has psychological problems with all of that—same with me. I know I am not alone. There are others who have the same problems. This experience by itself is very helpful…* [P5-FG4]

Although patients generally saw value in exchanging experiences with their peers, they also voiced concerns regarding reliability and credibility of information exchanged. A few HCPs responded positively to communicate with other HCPs regarding a patient's medication regime (ie, to resolve discrepancies) through the platform, particularly during transitions of care. However, the majority of HCPs were reluctant to communicate electronically with other providers and, especially not with patients, and feared an increased workload. Furthermore, they emphasized it should not be used for urgent matters as it would be impossible to answer requests in a timely manner:

> *...I imagine regular consultation hours. I don't read emails or do chats. I have no time for that during consultation hours; I must say clearly. In our practice, we have even stopped interruptions by telephone except in real emergencies...* [GP4-FG3]

Usability of the Medication Platform

Participants' expectations regarding usability of the medication platform were grouped into 2 subcategories: (1) user interface and (2) user-centered provision of information (see Table 4 and Multimedia Appendix 3). Overall, patients and HCPs expressed similar expectations on the issue of usability.

User Interface

Patient and HCP groups emphasized that medication information should be displayed in a clear and logical manner in a large font size on the platform. For instance, participants suggested listing medications in a chronological order, organizing them according to diagnosis (eg, T2DM, cardiovascular), and categorizing medications in relation to their short-term (eg, an antibiotic) or long-term use (eg, OHA):

> *The medication platform should be simply structured, there should be the long-term medications, then the on-demand medications or short-term medications, and easily accessible and changeable…* [GP3-FG1]

Overall, to support a patient's self-management, participants highlighted that the platform should be easy to navigate, to enable quick access to relevant medication information, and to observe a patient's current medication at a glance. After log in, the current medication list should be prominently displayed on the front page, thus enabling a quick overview of a patient's current medication. Likewise, participants commented it was important that the platform provided sufficient information to make informed decisions but at the same time was not overloaded with information:

> *However, one should not overload such a system, because otherwise you don't find anything and have to search…* [GP4-FG3]

Consequently, participants requested the platform should enable users to customize information density. For instance, some participants suggested to provide access to detailed information (eg, regarding side effects) via a link on the specific medication. Especially HCPs groups expected the platform to be tailored to suit their practice workflow and to function quickly and reliably.

User-Centered Provision of Information

From the participants' point of view, a user-centered provision of information in the platform was fundamental for the adoption and successful use by patients and HCPs. Most patients described their difficulties understanding medication information in consultations and in written information leaflets. Accordingly, patients as well as HCPs stressed that information should be evidence-based and provided in "nontechnical jargon" [P6-FG4] in the platform. In addition, the group of Turkish patients remarked it was important that the platform provided access to multilingual information. Both groups thought it would be beneficial to provide a glossary or encyclopedia to support comprehensibility of medical terms and a wiki for answering important and frequently asked questions:

> *Yes, common misunderstandings, frequently asked questions and common medication intake errors…* [HCA4-FG2]

Especially the patient focus groups highlighted they would benefit if visual aids (eg, icons, pictograms, images of medications, daily injection plan) and videos (eg, instruction video) were included in the platform to assist identifying, understanding, and using their medications appropriately:

> *When it comes to injection technique, a video would be very helpful…* [P2-FG1]

HCPs also emphasized to integrate visualization methods into the platform:

> *It would be great if you click on a medication a video is shown or something else…* [GP2-FG2]

Besides, patients' desired prior instruction and training for using the new system and some requested a toll free hotline and a user guide to aid navigation. Furthermore, patients wished the platform linked them to an expert helpline for personal medication counseling and further self-management support (eg, regional patient support groups).

Discussion

Principal Findings

The qualitative UCD approach enabled a deeper understanding of the requirements of patients with T2DM, GPs, and HCAs regarding functionalities and usability of a shared patient-centered, Web-based medication platform. Identifying key users' requirements early on is a critical step in the development and implementation of a new system to successfully support medication management and treatment of patients with T2DM and their HCPs.

In patients' view, a medication platform offers potential to improve their understanding, address their medication-related concerns, and support their medication self-management activities. HCPs, in turn, focused on the platforms' ability to aid comprehensive medication history taking and reconciliation across health care settings. Appropriate security and access methods, supported data entry, printing and sending information electronically, and tracking medication history were perceived by participants as essential functionalities. Although patients wanted automatic interaction checks and safety alerts, HCPs, on the contrary, were concerned that unspecific alerts confuse patients and lead to nonadherence. Furthermore, HCPs were opposed to patients' ability to withhold or restrict access to information in the platform. To optimize usability, there was consensus among participants to display information in a structured, chronological format, to provide information in lay language, to use visual aids and customize information content, and align the platform to users' workflow.

Most participating patients had safety concerns, and prior research suggested that patients most commonly avoid taking their medications due to concerns about adverse effects [5]. To address these concerns, patients desired comprehensive risk-related information, automatic interaction checks, and safety alerts in the platform. Undeniably, patients play a central role in managing medication-related risks, and need to be engaged in self-monitoring to improve medication use [41]. Keeping in mind that patients with T2DM often have complex medication regimens and receive pharmaceutical treatment from different providers, tailored and clinically meaningful safety alerts combined with clear instructions on how to proceed could facilitate early detection and reduce serious complications. On the other hand, some patients may not want or feel capable to use safety-related IT apps. Thus, it will be necessary to customize safety alerts to the specific needs of individual patients, for instance, to allow triggering the interaction check manually (ie, non-interruptive) or showing only high severity alerts. To date, however, lack of specificity and low sensitivity of medication alerts in clinical decision support systems is still a problem [42,43]. Comparable with HCPs' concerns, unmodified medication safety alerts without concurrent physician interpretation may create confusion and anxiety among patients and thus impede medication-taking. At this point, it is important to emphasize that the medication platform does not replace or substitute for patient-provider consultation but has the potential to complement instructions and self-management support given. Furthermore, the platform offers potential to transmit essential information (eg, access to complete medication regimen, diagnoses) among members of the health care team. This may, for instance, promote stronger physician-pharmacist collaboration to improve medication therapy and safety. Overall, however, HCPs had conflicting views regarding the provision of risk-related information in the platform. Most HCPs held negative attitudes toward safety alerts for patients and also anticipated an increase in their workload. Nevertheless, they perceived patients should be made aware of the most relevant adverse effects. Delbanco and colleagues [44] found that patients who had electronic access to care providers' notes felt more in control of their care, and reported improved medication adherence and minimal concerns without increasing providers' workload. Hence, unintended consequences to patients need to be explored further in the user-driven design process to see if the platform and specifically safety alerts are both beneficial and acceptable to patients and HCPs and exceed potential risks [43]. Perhaps patients may have a more sensible approach to safety alerts than HCPs fear.

Patients' ability to control and restrict access to medication-related information or change physicians prescribed medication evoked great concerns (ie, medicolegal liability) among HCPs and was perceived to threaten physicians control and the quality of care. Thus, patients' desired ownership over the platform seems to clash with the predominant approach held by HCPs. Similar to previous research [45], physicians were concerned to make suboptimal decisions about a patient's treatment due to incomplete or inaccurate information. In contrast, Haverhals et al [26] also concluded that health apps should provide patients the ability to selectively disclose information (eg, alternative medications) to different HCPs. Further research and ongoing involvement of intended users is therefore necessary to elaborate how to place control of the medication platform in the hands of patients while accounting for HCPs' needs. Moreover, introducing such a patient-held information system challenges current structures and requires a shift in patients' and HCPs' roles and responsibilities.

One of the greatest concerns among patients and HCPs were privacy and security issues, which seems more prominent in Germany than in many other countries, and thus delay a national e-medication initiative [31]. Security and privacy concerns, however, have also been identified previously as a potentially large barrier to personal health record use [16,46]. Further investigation of privacy, security, and legal concerns is needed to better understand what prompts these concerns by patients and HCPs and to ensure that users' concerns are adequately addressed. Moreover, ways of granting secure and effective emergency access to the medication platform need to be explored in the iterative development process.

Participants were enthusiastic about the platforms' potential to collect and store a patient's complete medication history in a structured format by engaging both patients and HCPs. Indeed,

medication errors are common in primary care and there is a need for better monitoring, patient education, and improved communication between patients and their HCPs [8,47]. Although participants positively viewed patients' ability to contribute their self-medication and symptoms (eg, ADEs, intolerances) to the platform, some HCPs also mentioned concerns regarding reliability and accurateness of patient-entered data. Tang and colleagues [16] emphasized that the reliability of data entered by patients depends on the nature of information per se, the patient's literacy level, and the motivations for recording the data. Providing data entry functionalities in the platform, for example, scanning the medication barcodes with a mobile device and integrating a plausibility checker, may help to improve information accuracy. Moreover, effective training of patients and HCPs will be a prerequisite to unfold the platform's full potential. Overall, facilitating patient engagement in medication history collection seems a promising approach to improve medication reconciliation, patient-provider communication, and thus patient safety [48].

Although the platform may offer new ways to mediate communication among patients and HCPs, this function received only little attention in the focus group discussions and was subject to concerns. Although patient medication reviews have also been identified previously as a valuable complementary source of information for patients [49], the reliability of this information has likewise been questioned [50]. Nevertheless, patient online communities have been suggested to facilitate patients' comprehension, informed decision making, and medication self-management [51]. Despite evidence that patients increasingly desire Web-based patient-provider communication [52], patients in this study did not request Web-based messaging with their HCPs. Maybe patients did not expect it to be an effective way to communicate with their HCPs about their medications or generally lacked ideas how the platform could facilitate patient-provider communication. Likewise, HCPs were reluctant to communicate electronically with patients, although a few HCPs saw potential for medication reconciliation through secure messaging with other HCPs.

Patients and HCPs had similar expectations regarding usability of the medication platform. For both groups, it was essential that the platform structured information chronologically in an intuitive, user-centered format (eg, customizable content in lay and multilingual language) on 1 screen and aligns to users' workflow. As also proposed earlier [16,30], special attention needs to be paid to health literacy issues when developing such a platform. Visual aids, pictograms, and customized videos requested by participants can enhance patients' understanding of how they should take their medications [53,54]. However, there are additional skills in terms of accessing and effectively using HIT, subsumed as eHealth literacy, that are required to fully engage with eHealth resources [55]. To date, patients and HCPs have not been trained sufficiently in the optimal use and implementation of HIT in medication management [56]. It is therefore vital that all essential aspects of using and implementing the medication platform should be included in the training, telephone, and on-site support [57]. Moreover, prior education of patients and HCPs on medication self-management (support) is vital to equip users with the essential skills and thus have implications for realizing the potential benefits of the platform. Further developments of the platform should also explore and incorporate features (eg, electronic diary, self-monitoring tools, nutrition module) that help patients to adopt and maintain a healthy lifestyle [2].

Strengths and Limitations

A strength of this qualitative study is that the in-depth perspectives of patients with T2DM who were diverse in age, duration of diabetes, and on a variety of medication regimes and their primary HCPs, including GPs and HCAs, were collected. By applying the principles of UCD, intended users have been involved early in the design and development process of the medication platform and thus increase the likelihood to achieve a functional and useful system. Although participants were purposefully selected and recruited from different health care settings, they may not be generalizable to the diabetes patient population or HCPs overall. As participants "opted in" to the focus groups, they may have greater interest in medication management and HIT, and may represent the perspectives of "early adopters," although computer experience differed among participants. Although this provides valuable insights on the needs of early adopters of HIT, we do not know the perspectives of potential participants who chose not to participate. Moreover, this study did not specifically focus on low-literate patients, their caregivers, or other HCP groups. Incorporating their perspectives may have generated a more extensive requirements elicitation. Despite the limitations, this study enables a comprehensive description of patients' and primary HCPs' requirements regarding a shared patient-centered, Web-based medication platform and will ultimately help to design the platform according to these needs.

Conclusions

The need to explore new approaches to facilitate medication management and treatment across health care sectors is an important issue that becomes increasingly important with the number of patients with T2DM. Given that patients are in control of their daily diabetes care [58] and the central users of the prospective system, their requirements need to be taken into account. This must, however, always be regarded in relation to the respective health literacy of each patient. No "one-size-fits-all" solution seems to be possible. The platform will be needed to be tailored to patients' needs and capabilities. Furthermore, reconciling differences in requirements of patients and HCPs, especially regarding medication safety alerts and access control, will necessitate substantial engagement by all intended user groups in the ongoing development process. Balancing patients and HCP's preferences is a prerequisite to empower patients and improve medication management and safety, while encouraging HCPs to use the platform. Once the prototype is developed, its evaluation will show how patients and HCPs evaluate and use the system, showing if the system has a good FITT and promotes the intended health outcomes.

Acknowledgments

First of all, the authors want to thank all study participants for sharing their perspectives in the course of this study. The authors also thank the Selbsthilfe Heidelberg and Mannheim (regional self-help office) as well as physicians and health care workers in the Rhine-Neckar region for helping to contact participants.

This paper presents independent research supported by grants from the German Federal Ministry of Education and Research (funding code 01KQ1003B). The contract between the sponsor and the Heidelberg University leaves the full responsibility for the scientific work, for the management of data, and for analysis and publication to the investigators.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Sample questions used in the focus groups with patients and healthcare professionals.

[PDF File (Adobe PDF File), 146KB - jmir_v20i3e105_app1.pdf]

Multimedia Appendix 2

Overview of required functionalities of the medication platform.

[PDF File (Adobe PDF File), 109KB - jmir_v20i3e105_app2.pdf]

Multimedia Appendix 3

Overview of requirements regarding usability of the medication platform.

[PDF File (Adobe PDF File), 80KB - jmir_v20i3e105_app3.pdf]

References

1. Guariguata L, Whiting DR, Hambleton I, Beagley J, Linnenkamp U, Shaw JE. Global estimates of diabetes prevalence for 2013 and projections for 2035. Diabetes Res Clin Pract 2014 Feb;103(2):137-149. [doi: 10.1016/j.diabres.2013.11.002] [Medline: 24630390]
2. World Health Organization. In: Sabaté E, editor. Adherence to Long-Term Therapies: Evidence for Action. Geneva: World Health Organization; 2003.
3. Norris SL, Engelgau MM, Narayan KM. Effectiveness of self-management training in type 2 diabetes: a systematic review of randomized controlled trials. Diabetes Care 2001 Mar;24(3):561-587. [Medline: 11289485]
4. Bailey SC, Oramasionwu CU, Wolf MS. Rethinking adherence: a health literacy-informed model of medication self-management. J Health Commun 2013 Dec;18(Suppl 1):20-30 [FREE Full text] [doi: 10.1080/10810730.2013.825672] [Medline: 24093342]
5. Brundisini F, Vanstone M, Hulan D, DeJean D, Giacomini M. Type 2 diabetes patients' and providers' differing perspectives on medication nonadherence: a qualitative meta-synthesis. BMC Health Serv Res 2015 Nov 23;15(1):516 [FREE Full text] [doi: 10.1186/s12913-015-1174-8] [Medline: 26596271]
6. Rubin RR. Adherence to pharmacologic therapy in patients with type 2 diabetes mellitus. Am J Med 2005 May;118(Suppl 5A):27S-34S. [doi: 10.1016/j.amjmed.2005.04.012] [Medline: 15850551]
7. Field TS, Mazor KM, Briesacher B, Debellis KR, Gurwitz JH. Adverse drug events resulting from patient errors in older adults. J Am Geriatr Soc 2007 Feb;55(2):271-276. [doi: 10.1111/j.1532-5415.2007.01047.x] [Medline: 17302666]
8. Sarkar U, Handley MA, Gupta R, Tang A, Murphy E, Seligman HK, et al. What happens between visits? Adverse and potential adverse events among a low-income, urban, ambulatory population with diabetes. Qual Saf Health Care 2010 Jun;19(3):223-228. [doi: 10.1136/qshc.2008.029116] [Medline: 20378619]
9. Bailey CJ, Kodack M. Patient adherence to medication requirements for therapy of type 2 diabetes. Int J Clin Pract 2011 Mar;65(3):314-322. [doi: 10.1111/j.1742-1241.2010.02544.x] [Medline: 21314869]
10. Krass I, Schieback P, Dhippayom T. Adherence to diabetes medication: a systematic review. Diabet Med 2015 Jun;32(6):725-737. [doi: 10.1111/dme.12651] [Medline: 25440507]
11. Ho PM, Rumsfeld JS, Masoudi FA, McClure DL, Plomondon ME, Steiner JF, et al. Effect of medication nonadherence on hospitalization and mortality among patients with diabetes mellitus. Arch Intern Med 2006 Sep 25;166(17):1836-1841. [doi: 10.1001/archinte.166.17.1836] [Medline: 17000939]
12. Krist AH, Woolf SH. A vision for patient-centered health information systems. JAMA 2011 Jan 19;305(3):300-301. [doi: 10.1001/jama.2010.2011] [Medline: 21245186]

13. Osborn CY, Mayberry LS, Mulvaney SA, Hess R. Patient web portals to improve diabetes outcomes: a systematic review. Curr Diab Rep 2010 Dec;10(6):422-435 [FREE Full text] [doi: 10.1007/s11892-010-0151-1] [Medline: 20890688]
14. Osborn CY, Mayberry LS, Wallston KA, Johnson KB, Elasy TA. Understanding patient portal use: implications for medication management. J Med Internet Res 2013;15(7):e133 [FREE Full text] [doi: 10.2196/jmir.2589] [Medline: 23823974]
15. Urowitz S, Wiljer D, Dupak K, Kuehner Z, Leonard K, Lovrics E, et al. Improving diabetes management with a patient portal: a qualitative study of diabetes self-management portal. J Med Internet Res 2012;14(6):e158 [FREE Full text] [doi: 10.2196/jmir.2265] [Medline: 23195925]
16. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc 2006 Mar;13(2):121-126 [FREE Full text] [doi: 10.1197/jamia.M2025] [Medline: 16357345]
17. Price M, Bellwood P, Kitson N, Davies I, Weber J, Lau F. Conditions potentially sensitive to a personal health record (PHR) intervention, a systematic review. BMC Med Inform Decis Mak 2015 Apr 18;15(1):32 [FREE Full text] [doi: 10.1186/s12911-015-0159-1] [Medline: 25927384]
18. Cresswell K, Sheikh A. Organizational issues in the implementation and adoption of health information technology innovations: an interpretative review. Int J Med Inform 2013 May;82(5):e73-e86. [doi: 10.1016/j.ijmedinf.2012.10.007] [Medline: 23146626]
19. Archer N, Fevrier-Thomas U, Lokker C, McKibbon KA, Straus SE. Personal health records: a scoping review. J Am Med Inform Assoc 2011;18(4):515-522 [FREE Full text] [doi: 10.1136/amiajnl-2011-000105] [Medline: 21672914]
20. Sligo J, Gauld R, Roberts V, Villa L. A literature review for large-scale health information system project planning, implementation and evaluation. Int J Med Inform 2017 Jan;97:86-97. [doi: 10.1016/j.ijmedinf.2016.09.007] [Medline: 27919399]
21. Vredenburg K, Isensee S, Righi C. User-Centered Design: An Integrated Approach. Upper Saddle River, NJ: Prentice Hall PTR; 2001.
22. Morales Rodriguez M, Casper G, Brennan PF. Patient-centered Design: The Potential of User-centered Design in Personal Health Records. J AHIMA 2007 Apr;78(4):44-6; quiz 49-50. [Medline: 17455846]
23. LeRouge C, Wickramasinghe N. A review of user-centered design for diabetes-related consumer health informatics technologies. J Diabetes Sci Technol 2013 Jul;7(4):1039-1056 [FREE Full text] [Medline: 23911188]
24. Ammenwerth E, Iller C, Mahler C. IT-adoption and the interaction of task, technology and individuals: a fit framework and a case study. BMC Med Inform Decis Mak 2006 Jan 09;6(1):3 [FREE Full text] [doi: 10.1186/1472-6947-6-3] [Medline: 16401336]
25. Bailey SC, Belter LT, Pandit AU, Carpenter DM, Carlos E, Wolf MS. The availability, functionality, and quality of mobile applications supporting medication self-management. J Am Med Inform Assoc 2014 May;21(3):542-546 [FREE Full text] [doi: 10.1136/amiajnl-2013-002232] [Medline: 24163156]
26. Haverhals LM, Lee CA, Siek KA, Darr CA, Linnebur SA, Ruscin JM, et al. Older adults with multi-morbidity: medication management processes and design implications for personal health applications. J Med Internet Res 2011 Jun;13(2):e44 [FREE Full text] [doi: 10.2196/jmir.1813] [Medline: 21715286]
27. Nelson LA, Bethune MC, Lagotte AE, Osborn CY. The Usability of Diabetes MAP: A Web-delivered Intervention for Improving Medication Adherence. JMIR Hum Factors 2016 May 12;3(1):e13 [FREE Full text] [doi: 10.2196/humanfactors.5177] [Medline: 27174496]
28. El-Gayar O, Timsina P, Nawar N, Eid W. Mobile applications for diabetes self-management: status and potential. J Diabetes Sci Technol 2013 Jan 01;7(1):247-262 [FREE Full text] [Medline: 23439183]
29. Ronda MC, Dijkhorst-Oei LT, Rutten GE. Reasons and barriers for using a patient portal: survey among patients with diabetes mellitus. J Med Internet Res 2014 Nov 25;16(11):e263 [FREE Full text] [doi: 10.2196/jmir.3457] [Medline: 25424228]
30. Irizarry T, DeVito Dabbs A, Curran CR. Patient Portals and Patient Engagement: A State of the Science Review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
31. Gall W, Aly AF, Sojer R, Spahni S, Ammenwerth E. The national e-medication approaches in Germany, Switzerland and Austria: A structured comparison. Int J Med Inform 2016 Sep;93:14-25. [doi: 10.1016/j.ijmedinf.2016.05.009] [Medline: 27435943]
32. Kitzinger J. Qualitative research. Introducing focus groups. BMJ 1995 Jul 29;311(7000):299-302 [FREE Full text] [Medline: 7633241]
33. Bruseberg A, McDonagh-Philp D. New product development by eliciting user experience and aspirations. Int J Hum Comput Stud 2001 Oct;55(4):435-452. [doi: 10.1006/ijhc.2001.0479]
34. INFOPAT. URL: http://www.infopat.eu/ [accessed 2017-08-07] [WebCite Cache ID 6sX7jrpO9]
35. Längst G, Seidling HM, Stützle M, Ose D, Baudendistel I, Szecsenyi J, et al. Factors associated with medication information in diabetes care: differences in perceptions between patients and health care professionals. Patient Prefer Adherence 2015 Oct;9:1431-1441 [FREE Full text] [doi: 10.2147/PPA.S88357] [Medline: 26508840]

36. Bernhard G, Ose D, Baudendistel I, Seidling HM, Stützle M, Szecsenyi J, et al. Understanding Challenges, Strategies, and the Role of Support Networks in Medication Self-management Among Patients With Type 2 Diabetes. Diabetes Educ 2017 Apr;43(2):190-205. [doi: 10.1177/0145721717697243] [Medline: 28340544]
37. Bowen GA. Naturalistic inquiry and the saturation concept: a research note. Qual Res 2008 Feb;8(1):137-152. [doi: 10.1177/1468794107085301]
38. Mayring P. Qualitative Inhaltsanalyse. Grundlagen und Techniken [Qualitative content analysis. Principles and techniques]. 11th edition. Weinheim & Basel: Beltz Verlag; Sep 2010.
39. Kuckartz U. Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung (Qualitative Text Analysis. Methods, Practice, Using Software). Weinheim & Basel: Beltz Juventa; Jul 2012.
40. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
41. Ryan R, Santesso N, Lowe D, Hill S, Grimshaw J, Prictor M, et al. Interventions to improve safe and effective medicines use by consumers: an overview of systematic reviews. Cochrane Database Syst Rev 2014 Apr 29(4):CD007768. [doi: 10.1002/14651858.CD007768.pub3] [Medline: 24777444]
42. Topaz M, Seger DL, Slight SP, Goss F, Lai K, Wickner PG, et al. Rising drug allergy alert overrides in electronic health records: an observational retrospective study of a decade of experience. J Am Med Inform Assoc 2016 May;23(3):601-608. [doi: 10.1093/jamia/ocv143] [Medline: 26578227]
43. Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. J Am Med Inform Assoc 2004 Mar;11(2):104-112 [FREE Full text] [doi: 10.1197/jamia.M1471] [Medline: 14633936]
44. Delbanco T, Walker J, Bell SK, Darer JD, Elmore JG, Farag N, et al. Inviting patients to read their doctors' notes: a quasi-experimental study and a look ahead. Ann Intern Med 2012 Oct 2;157(7):461-470 [FREE Full text] [doi: 10.7326/0003-4819-157-7-201210020-00002] [Medline: 23027317]
45. Tierney WM, Alpert SA, Byrket A, Caine K, Leventhal JC, Meslin EM, et al. Provider responses to patients controlling access to their electronic health records: a prospective cohort study in primary care. J Gen Intern Med 2015 Jan;30(Suppl 1):S31-S37 [FREE Full text] [doi: 10.1007/s11606-014-3053-0] [Medline: 25480720]
46. Kruse CS, Argueta DA, Lopez L, Nair A. Patient and provider attitudes toward the use of patient portals for the management of chronic disease: a systematic review. J Med Internet Res 2015 Feb 20;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]
47. Gandhi TK, Weingart SN, Borus J, Seger AC, Peterson J, Burdick E, et al. Adverse drug events in ambulatory care. N Engl J Med 2003 Apr 17;348(16):1556-1564. [doi: 10.1056/NEJMsa020703] [Medline: 12700376]
48. Schnipper JL, Gandhi TK, Wald JS, Grant RW, Poon EG, Volk LA, et al. Design and implementation of a web-based patient portal linked to an electronic health record designed to improve medication safety: the Patient Gateway medications module. Inform Prim Care 2008;16(2):147-155 [FREE Full text] [Medline: 18713531]
49. Hughes S, Cohen D. Can online consumers contribute to drug knowledge? A mixed-methods comparison of consumer-generated and professionally controlled psychotropic medication information on the internet. J Med Internet Res 2011 Jul 29;13(3):e53 [FREE Full text] [doi: 10.2196/jmir.1716] [Medline: 21807607]
50. Adusumalli S, Lee H, Hoi Q, Koo S, Tan IB, Ng PC. Assessment of Web-Based Consumer Reviews as a Resource for Drug Performance. J Med Internet Res 2015 Aug 28;17(8):e211 [FREE Full text] [doi: 10.2196/jmir.4396] [Medline: 26319108]
51. Wicks P, Massagli M, Frost J, Brownstein C, Okun S, Vaughan T, et al. Sharing health data for better outcomes on PatientsLikeMe. J Med Internet Res 2010 Jun 14;12(2):e19 [FREE Full text] [doi: 10.2196/jmir.1549] [Medline: 20542858]
52. Lin CT, Wittevrongel L, Moore L, Beaty BL, Ross SE. An Internet-based patient-provider communication system: randomized controlled trial. J Med Internet Res 2005 Aug 05;7(4):e47 [FREE Full text] [doi: 10.2196/jmir.7.4.e47] [Medline: 16236699]
53. Katz MG, Kripalani S, Weiss BD. Use of pictorial aids in medication instructions: a review of the literature. Am J Health Syst Pharm 2006 Dec 01;63(23):2391-2397. [doi: 10.2146/ajhp060162] [Medline: 17106013]
54. Davis SA, Carpenter D, Cummings DM, Lee C, Blalock SJ, Scott JE, et al. Patient adoption of an internet based diabetes medication tool to improve adherence: A pilot study. Patient Educ Couns 2017 Jan;100(1):174-178. [doi: 10.1016/j.pec.2016.07.024] [Medline: 27480530]
55. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
56. van Stiphout F, Zwart-van Rijkom JE, Maggio LA, Aarts JE, Bates DW, van Gelder T, et al. Task analysis of information technology-mediated medication management in outpatient care. Br J Clin Pharmacol 2015 Sep;80(3):415-424 [FREE Full text] [doi: 10.1111/bcp.12625] [Medline: 25753467]
57. Marien S, Krug B, Spinewine A. Electronic tools to support medication reconciliation: a systematic review. J Am Med Inform Assoc 2017 Jan;24(1):227-240. [doi: 10.1093/jamia/ocw068] [Medline: 27301747]
58. Lange K. Supporting the active participation of patients with diabetes type 2: Personal beliefs, risk management and behaviour modification. Diabetologe 2006 Sep;2:484-492 [FREE Full text] [doi: 10.1007/s11428-006-0077-z]

Abbreviations

ADE: adverse drug event
DMP: disease management program
FITT: Fit between Individuals, Task, and Technology
GP: general practitioner
HCA: health care assistant
HCP: health care professional
HIT: health information technology
Infopat Rhein-Neckar: Information Technologies for Patient-Centered Health Care
OHA: oral hypoglycemic agents
OTC: over-the-counter medications
PC: personal computer
T2DM: type 2 diabetes mellitus
UCD: user-centered design

Edited by G Eysenbach; submitted 23.08.17; peer-reviewed by S Ross, S Hussain, C Kruse, S Manzanero; comments to author 05.10.17; revised version received 24.11.17; accepted 07.12.17; published 27.03.18.

Please cite as:
Bernhard G, Mahler C, Seidling HM, Stützle M, Ose D, Baudendistel I, Wensing M, Szecsenyi J
Developing a Shared Patient-Centered, Web-Based Medication Platform for Type 2 Diabetes Patients and Their Health Care Providers: Qualitative Study on User Requirements
J Med Internet Res 2018;20(3):e105
URL: http://www.jmir.org/2018/3/e105/
doi:10.2196/jmir.8666
PMID:29588269

Original Paper

Rapid and Accurate Behavioral Health Diagnostic Screening: Initial Validation Study of a Web-Based, Self-Report Tool (the SAGE-SR)

Benjamin Brodey[1], MPH, MD; Susan E Purcell[1], PhD; Karen Rhea[2], MD; Philip Maier[1], BS; Michael First[3], MD; Lisa Zweede[1], MD; Manuela Sinisterra[1], BA; M Brad Nunn[2], PhD; Marie-Paule Austin[4], MD, FRANZCP; Inger S Brodey[5], PhD

[1]TeleSage, Chapel Hill, NC, United States

[2]Centerstone, Nashville, TN, United States

[3]Department of Psychiatry, Columbia University, New York City, NY, United States

[4]School of Psychiatry, University of New South Wales, Sydney, Australia

[5]Department of English and Comparative Literature, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Corresponding Author:
Benjamin Brodey, MPH, MD
TeleSage
201 East Rosemary Street
Chapel Hill, NC, 27514
United States
Phone: 1 919 942 8849
Email: bb@telesage.com

Abstract

Background: The Structured Clinical Interview for DSM (SCID) is considered the gold standard assessment for accurate, reliable psychiatric diagnoses; however, because of its length, complexity, and training required, the SCID is rarely used outside of research.

Objective: This paper aims to describe the development and initial validation of a Web-based, self-report screening instrument (the Screening Assessment for Guiding Evaluation-Self-Report, SAGE-SR) based on the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) and the SCID-5-Clinician Version (CV) intended to make accurate, broad-based behavioral health diagnostic screening more accessible within clinical care.

Methods: First, study staff drafted approximately 1200 self-report items representing individual granular symptoms in the diagnostic criteria for the 8 primary SCID-CV modules. An expert panel iteratively reviewed, critiqued, and revised items. The resulting items were iteratively administered and revised through 3 rounds of cognitive interviewing with community mental health center participants. In the first 2 rounds, the SCID was also administered to participants to directly compare their Likert self-report and SCID responses. A second expert panel evaluated the final pool of items from cognitive interviewing and criteria in the DSM-5 to construct the SAGE-SR, a computerized adaptive instrument that uses branching logic from a screener section to administer appropriate follow-up questions to refine the differential diagnoses. The SAGE-SR was administered to healthy controls and outpatient mental health clinic clients to assess test duration and test-retest reliability. Cutoff scores for screening into follow-up diagnostic sections and criteria for inclusion of diagnoses in the differential diagnosis were evaluated.

Results: The expert panel reduced the initial 1200 test items to 664 items that panel members agreed collectively represented the SCID items from the 8 targeted modules and DSM criteria for the covered diagnoses. These 664 items were iteratively submitted to 3 rounds of cognitive interviewing with 50 community mental health center participants; the expert panel reviewed session summaries and agreed on a final set of 661 clear and concise self-report items representing the desired criteria in the DSM-5. The SAGE-SR constructed from this item pool took an average of 14 min to complete in a nonclinical sample versus 24 min in a clinical sample. Responses to individual items can be combined to generate DSM criteria endorsements and differential diagnoses, as well as provide indices of individual symptom severity. Preliminary measures of test-retest reliability in a small, nonclinical sample were promising, with good to excellent reliability for screener items in 11 of 13 diagnostic screening modules

(intraclass correlation coefficient [ICC] or kappa coefficients ranging from .60 to .90), with mania achieving fair test-retest reliability (ICC=.50) and other substance use endorsed too infrequently for analysis.

Conclusions: The SAGE-SR is a computerized adaptive self-report instrument designed to provide rigorous differential diagnostic information to clinicians.

(J Med Internet Res 2018;20(3):e108) doi:10.2196/jmir.9428

KEYWORDS

mental health; differential diagnosis; surveys and questionnaires; self-report; primary health care; computer-assisted diagnosis

Introduction

The Structured Clinical Interview for DSM-5 (SCID-5) is currently accepted as the gold standard in psychiatric diagnosis and is regularly used in research settings where the accurate diagnosis of primary and comorbid disorders is required for the appropriate determination of study eligibility and assignment to a research condition [1-3]. The SCID is also frequently used as the standard against which other diagnostic instruments are validated (eg, [4-8]). The structured format of the SCID with its direct adherence to Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria accounts for its strong test-retest and inter-rater reliability for most diagnoses [1,2,3,9]. Overall, the full SCID-5-Research Version (RV) covers 63 diagnoses, takes an average of 90 min to administer, and requires considerable clinician training [2,10]. The Clinician Version (CV) of the SCID for DSM-5 (SCID-5-CV), released in 2014, consists of 10 modules that cover 39 of the most common diagnoses seen in clinical practice and allows screening for an additional 16 diagnoses [1]. Although it is easy to select individual SCID modules for administration, more complex customizations of items and diagnoses within modules can be difficult to implement.

To streamline use of the SCID in research and to make it more accessible for use in clinical settings, clinician-administered, Web-based versions of the SCID instruments were developed including the NetSCID-5-Clinician Version (NetSCID-5-CV), which covers the same disorders as the SCID-5-CV paper version; the NetSCID-5-Research Version (NetSCID-5-RV), which covers the same diagnostic modules of the paper version of the SCID-5-RV; and the NetSCID-5-Personality Disorder (PD) Version, which covers the 10 DSM-5 PDs across Clusters A, B, and C, as well as other specified PD [11]. In a validation study versus the paper version of the SCID-IV-RV, an earlier version of the NetSCID-RV demonstrated fewer data entry and branching errors than the paper version, was preferred by clinicians over the paper version, and was easier to administer [11]. In addition, anecdotal reports from clinicians indicate that NetSCID administration requires 30% less time than the corresponding paper SCID [11]. However, despite its advantages over the paper version of the SCID, the clinician NetSCID administration still requires significant clinician time and training, which may pose too large a burden for routine clinical care settings [12] and in epidemiological studies evaluating large numbers of participants where clinician-based interviewing becomes logistically prohibitive.

Perhaps primarily because of the need for time-efficient diagnostic practices, routine clinical practice continues to rely predominantly on unstructured clinical interviews [13], despite mounting evidence that doing so often results in missed comorbidities [14-16], missed diagnoses [17], and less-specific diagnoses (eg, adjustment disorders vs more specific mood or anxiety disorders) [15]. Some researchers suggest that clinicians who do not use structured interviews may sometimes narrow their diagnostic focus too quickly, thereby missing comorbid diagnoses, whereas structured interviews ensure clinicians assess a broader range of clinical diagnoses [3]. Research has shown that accurate diagnosis has implications for clients' engagement in treatment and treatment outcomes [18], possibly linked to the role accurate diagnosis plays in the appropriate selection of evidence-based treatments [19]. The need for time-efficient and rigorous diagnostic practices is probably highest in primary care, where behavioral health problems are common presenting complaints and clinicians are under the highest pressure to assess and treat patients in a time-efficient manner [20].

Given the tension between the need for accurate diagnosis and the limited resource of clinician time in routine clinical practice, especially in primary care, one proposed solution is for patients to take a self-administered diagnostic screening questionnaire before their intake interview with a clinician. Results of this self-report screening measure could focus the clinician's diagnostic expertise on a differential diagnosis during the face-to-face client interview [10,12,21], as well as satisfy the current mandate by the Affordable Care Act to routinely screen patients in primary care for depression and alcohol abuse [20]. To ensure that clinicians do not miss potential comorbidities, such screening questionnaires should be broad-based and cover a wide range of diagnoses frequently seen in clinical practice. Most currently available broad-based diagnostic screening measures either require a clinician or other trained interviewer to administer them (eg, Mini International Neuropsychiatric Interview–Clinician Rated, MINI-CR [7]; World Health Organization World Mental Health Composite International Diagnostic Interview, WHO WMH-CIDI [6,22]) or are only available with paper and pencil administration and scoring or do not correspond directly to DSM-5 criteria (eg, Clinical Interview Schedule–Revised [23-25]; Mini International Neuropsychiatric Interview–Patient Rated, MINI-PR [7]; the Psychiatric Diagnostic Screening Questionnaire [8]; Primary Care Evaluation of Mental Disorders [26]).

The goal of this study was to develop a computerized adaptive self-report assessment based on the SCID and DSM-5 criteria [27] complete with self-scoring and instantaneous report generation of a rigorous differential diagnosis for clinicians. Ideally, these reports would be immediately accessible through the client's electronic health record. As with the NetSCID [11],

the resulting assessment would be a HIPAA-compliant, Web-based software program that patients could complete at a mental health clinic or a primary care clinic using a desktop computer, laptop, tablet, or smartphone. Reports would enable clinicians to initiate a more focused routine diagnostic interview based on considerable background knowledge of the patient's symptoms.

Methods

Stage I: Self-Report Item Pool Development

As a first step, we authored a set of approximately 1200 unique self-report items that mirrored the questions in the SCID for DSM-IV and corresponded with criteria outlined in the DSM-IV-TR. In anticipation of the release of DSM-5, we also developed items intended to represent the few anticipated changes to diagnostic criteria occurring between DSM-IV and DSM-5 (prospective changes were made available online before the DSM-5's publication date). TeleSage staff developed these items using a rigorous methodology first developed and successfully implemented in our previous instrument development work [28]. Self-report items were drafted for 13 diagnostic categories judged to be the most commonly encountered in clinical practice by the developers of the SCID-CV [1]: (1) depressive disorders, (2) manic and hypomanic disorders, (3) generalized anxiety disorder (GAD), (4) panic disorder, (5) agoraphobia, (6) social anxiety disorder, (7) obsessive-compulsive disorder (OCD), (8) posttraumatic stress disorder (PTSD), (9) adult attention-deficit/hyperactivity disorder (ADHD), (10) psychotic disorders, (11) alcohol use disorder, (12) cannabis use disorder, and (13) other substance use disorders. Whenever the DSM included differing symptoms for "adolescents," that wording was included as well to maximize the utility and flexibility of the resulting instrument.

During the item development process, staff members strove to keep items very simple by developing items that omitted lead phrases; omitted contingencies; included only a single concept; omitted idiomatic language; adhered to a 5-point Likert scale (never, rarely, sometimes, often, and always) wherever possible; used simple English language words; simple syntax, so as not to exceed a fifth grade reading level; and a consistent timeframe, where applicable, depending on the DSM-specified timeframe. These strategies were aimed at producing items that were easy to read, easy to understand, could be understood by non-native English speakers, and were amenable to direct translation into other languages. For SCID questions that were not straightforward (eg, questions that had multiple components), several simple self-report items were created. For example, to represent depression criteria 1A "In the last month, has there been a time when you were feeling hopeless, depressed, or down most of the day nearly every day," 5 items were drafted: (1) I felt sad; (2) I felt depressed; (3) I felt irritable; (4) I felt hopeless; and a fifth item relating to "most of the day nearly every day." We did not use the term "down" as it is idiomatic. We did create an item for hopelessness as it is in the DSM-5, although it is absent in DSM-IV-TR. We also created an item for irritability as it is a criterion for youth.

Given that we intended to develop readily understandable, clear expressions of clinical symptoms in simple language, we acknowledge that the content of some of our items may overlap with other existing measures. Indeed, 3 of the 5 items just referenced to represent DSM-5 depression criteria 1A are also present in the Patient-Reported Outcomes Measurement Information System (PROMIS) [29] item bank for depression. In developing their emotional distress items (within the domains of depression, anger, and anxiety), the PROMIS researchers, using an item response theory process, identified 78 different depression scales in the literature and found considerable overlap in the items covered [30]. In discussing the intellectual property rights regarding such items, these researchers noted that this overlap likely existed because the items "reflected generic aspects of emotional distress and the everyday language in which it is described" and thus "regarded them as part of the public domain because they reflected common-sense ideas about emotional distress" [30]. Our consultations with 2 lawyers specializing in intellectual property issues also supported the perspective that this is the case for simply worded individual items and small groups of items.

The completed item pool was iteratively reviewed by a panel of 7 experts, including Michael First, MD, the primary author of the SCID, 2 other psychiatrists, and 4 psychologists with combined expertise in community mental health, SCID items and administration, and mental health item development. Items were presented in tables populated with the original wording of the DSM criteria, the corresponding SCID item wording, and the proposed self-report items. The expert panel rated the clarity (1=unclear, 2=needs revision, and 3=clear) and correspondence with DSM-defined criteria (1=does not sufficiently correspond to DSM criteria, 2=needs to be rewritten to fully correspond to DSM criteria, and 3=directly and fully corresponds to DSM criteria) of each self-report item. Panel members also identified any missing concepts, offered suggestions for item rewrites, and discussed revised items by email and phone until consensus was reached on a final pool of items.

Stage II: Cognitive Interviewing

The self-report item pool was divided into 2, with 6 to 7 diagnostic categories (approximately 4 SCID modules) in each half. After engaging in an institutional review board (IRB)-approved informed consent process, participants were given the half of the item pool that corresponded with their individual chart diagnosis. Both halves were then tested and revised over 3 rounds of cognitive interviewing (CI). After each round of CI, session summaries were analyzed by TeleSage staff. All items that posed difficulty for 20% or more of the participants were either omitted or rewritten for the next round of CI.

CI is a scientific technique that uses *verbal probes* and *verbal think alouds* to determine the perceived meaning of survey questions [31]. For this study, the cognitive interviewer presented each participant with a block of self-report items that corresponded to a single diagnostic category at a time. Item sets pertaining to each diagnostic category were presented in a balanced, randomized order to control for order effects and ensure that majority of the questions were completed.

After reading an item aloud, participants marked their responses to the items. In addition, participants were instructed to circle any item they perceived as unclear or confusing as they completed the self-report assessment. Participants were also encouraged to think aloud while they answered each item. After participants completed the self-report items, the cognitive interviewer asked follow-up questions to further assess the reason the participant found each circled item unclear or confusing, while also confirming that the participant understood the meaning and intent of items that were not circled. For example, the cognitive interviewer would point out specific words in the question and ask for the meaning of that word (eg, "Can you tell me what *irritable* means to you?") or ask, for example, behaviors (eg, "You indicated that you "often" feel sad. Can you give me some examples of how you have felt sad in the past two weeks?"). This process continued until the interviewer probed all items. Interviews were recorded on a digital recorder, and the cognitive interviewer took objective, not interpretive, notes during the session pertaining to the participant's responses as well. After the interview, the cognitive interviewer listened to the audio file as needed and converted the notes from the session into a summary indicating items that were particularly difficult for the participant to answer or caused confusion, and items for which the participant's interpretation did not reflect the item's intent. By having participants describe all their thoughts out loud as they work their way through questions, it is possible to identify many of the potential problems that could affect a patient's response in unintended ways. Using CI to hone questions should improve the likelihood that individual items will ultimately have good psychometric characteristics during quantitative validation.

Each of the 3 rounds of CI was conducted with unique participants who engaged in an individual interview; no participant was interviewed twice. Participants in the first 2 rounds of CI were also given a clinician-administered SCID. This SCID contained the same modules (diagnostic categories) that the participants completed in the self-report item pool and included the participant's specific chart diagnosis. To account for any learning effect, participants were randomized so that half of the participants took the SCID first and half completed the self-report items and CI first.

Stage III: Screening Assessment for Guiding Evaluation-Self-Report Instrument Construction and Initial Validation

An expert panel was convened for this next stage to convert the self-report item pool into the computerized adaptive Screening Assessment for Guiding Evaluation-Self-Report (SAGE-SR). The panel included 2 psychiatrists, 2 clinical psychologists, 1 physician, TeleSage staff members with backgrounds in psychology as well as expertise in mental health item development and SCID administration, and TeleSage staff computer programmers with expertise in computer-adaptive instrument development. To construct an easily understood instrument that could be administered in a time-efficient manner, the SAGE-SR was constructed to have an initial 65-question screener which covered the same 13 diagnostic categories for which items were drafted in stage I. Respondents would need to endorse screener items at a sufficient threshold (set by the expert panel) within each diagnosis to "screen in" and branch to the remaining self-report items necessary to determine if respondents meet criteria for that diagnosis to be included in the final differential diagnosis. Possible diagnoses that could be returned in this differential diagnosis are presented in Table 1, along with the corresponding representation of diagnoses in the SCID-5-CV.

The expert panel examined the newly released DSM-5 criteria for each of the diagnoses covered by the self-report items to determine the most appropriate items for inclusion on the screener using clinical judgment for best fit and criteria that were "essential" or central to each diagnosis. For example, to meet DSM-5 criteria for major depressive disorder, 5 or more of a series of 9 symptoms must be present during the same 2-week period and represent a change from previous functioning [27]; however, 1 of these 5 symptoms must be either depressed mood or loss of interest or pleasure. Thus, the expert panel selected 3 self-report items for the screener to represent depressed mood ("I felt sad," "I felt depressed," and "I felt hopeless") and 3 self-report items for the screener to represent loss of interest or pleasure ("I enjoyed life"—reverse coded, "I had difficulty enjoying things that I used to enjoy," and "I was interested in my usual activities"—reverse coded). If a respondent met the threshold set by the expert panel on these screener items, the adaptive SAGE-SR would present the remaining depressive disorder items after the respondent completed the screener to determine if the respondent endorsed sufficient criteria for any depressive disorder to be considered for differential diagnosis. The expert panel also set the thresholds for determining whether respondents had endorsed sufficient criteria between the screener and follow-up questions for diagnoses to be reported for clinician consideration for differential diagnosis.

Once the initial instrument was constructed and programmed for Web-based administration (via personal computer, tablet, or smartphone), TeleSage staff members piloted and tested the Web-based administration of the SAGE-SR to identify any programming glitches. Following this process, healthy participants were recruited to take the SAGE-SR for the purpose of measuring administration time, assessing the appropriateness of the thresholds for screening and differential diagnosis set by the expert panel, identifying any remaining areas of confusion regarding item administration, and for preliminary quantitative validation. A subset of these participants returned for a second session within 1 week for the purpose of assessing test-retest reliability and how consistently participants screened into follow-up sections and received diagnoses for differential diagnostic consideration. All participants underwent a full informed consent process before engaging in any study procedures; all study and consent procedures were IRB-approved before the commencement of participant enrollment.

Table 1. Comparison of diagnoses covered by Screening Assessment for Guiding Evaluation-Self-Report and the Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Clinician Version.

DSM-5[a] diagnostic category	SAGE-SR[b] diagnoses or episodes covered	SCID-5-CV[c] diagnoses or episodes covered[d]
Mood disorders		
Major depressive episode	✓[e]	✓[f]
Manic episode	✓[e]	✓[f]
Hypomanic episode	✓[e]	✓[f]
Persistent depressive disorder	✓	✓
Major depressive disorder	✓	✓
Other specified depressive disorder	✓	✓
Bipolar I disorder	✓	✓
Bipolar II disorder	✓	✓
Other specified bipolar disorder	✓	✓
Anxiety disorders		
Panic disorder	✓	✓
Agoraphobia	✓	✓
Social anxiety disorder	✓	✓
Generalized anxiety disorder	✓	✓
Other specified anxiety disorder		✓
Anxiety disorder due to another medical condition		✓
Substance/medication-induced anxiety disorder		✓
Obsessive-compulsive and related disorders		
Obsessive-compulsive disorder	✓	✓
Obsessive-compulsive and related disorder due to another medical condition		✓
Substance/medication-induced obsessive-compulsive and related disorder		✓
Trauma- and stressor-related disorders		
Posttraumatic stress disorder	✓	✓
Neurodevelopmental disorders		
Attention-deficit/hyperactivity disorder	✓	✓
Psychotic disorders		
Schizophrenia	✓	✓
Schizophreniform disorder	✓	✓
Schizoaffective disorder	✓	✓
Delusional disorder	✓	✓
Brief psychotic disorder	✓	✓
Other specified psychotic disorder	✓	✓
Psychotic disorder due to another medical condition		✓
Substance/medication-induced psychotic disorder		✓
Substance-related and addictive disorders		
Alcohol use disorder	✓	✓
Cannabis use disorder	✓	✓
Inhalant use disorder	✓	✓
Other hallucinogen use disorder	✓	✓

DSM-5[a] diagnostic category	SAGE-SR[b] diagnoses or episodes covered	SCID-5-CV[c] diagnoses or episodes covered[d]
Opioid use disorder	✓	✓
Phencyclidine use disorder	✓	✓
Sedative, hypnotic, or anxiolytic use disorder	✓	✓
Stimulant use disorder	✓	✓
Other or unknown substance use disorder	✓	✓
Adjustment disorders		✓

[a]DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.

[b]SAGE-SR: Screening Assessment for Guiding Evaluation-Self-Report.

[c]SCID-5-CV: Clinician Version (CV) of the SCID for DSM-5.

[d]Screening questions are available on the SCID-5-CV for the following additional disorders: specific phobia, separation anxiety disorder, hoarding disorder, body dysmorphic disorder, trichotillomania, excoriation disorder, insomnia disorder, hypersomnolence disorder, anorexia nervosa, bulimia nervosa, binge eating disorder, avoidant/restrictive food intake disorder, somatic symptom disorder, illness anxiety disorder, intermittent explosive disorder, gambling disorder.

[e]Current episodes covered; past episodes under development.

[f]Current and past episodes covered.

After completing the SAGE-SR, participants were asked to provide feedback on whether they found any items confusing or unclear, whether they found items repetitive, whether the instructions were clear, what they thought about the length of the assessment, how well the progress bar and radio buttons on the device functioned, as well as any other comments they wanted to offer on what worked and what did not.

In addition, an initial sample of respondents was recruited from an outpatient public sector mental health clinic based in Tennessee; initial results from this population were used primarily to examine administration time in a clinical population. Further recruitment at this site is ongoing and will be used for future quantitative validation in a clinical population.

Results

Stage I: Item Development

The expert panel iteratively rated, discussed, and rewrote items until each item scored on average 2.5 or higher for clarity and correspondence with its respective DSM symptom. A consensus was reached on a final pool of 664 items that expert panel members agreed were clear, easy to understand, and collectively represented the items from 8 SCID-IV-CV modules, thus covering the DSM-IV-TR criteria for each of 13 diagnostic categories (see Table 1 for diagnostic coverage).

Stage II: Cognitive Interviewing

A total of 50 adult community mental health outpatients, including individuals with severe and persistent mental illness, were recruited from 2 locations at Centerstone, a private nonprofit mental health organization, in Nashville, TN, and in Bloomington, IN. Participants were recruited to ensure that they (according to their chart diagnoses) represented all 13 diagnostic categories in the self-report items (or 8 SCID-5-CV modules); participants ranged in age from 18 to 68 years (mean 39.9) and were 60% female (30/50), 86% white (43/50), 12% African American (6/50), and 2% Native American (1/50).

For the first round of CI, a total of 18 participants responded to approximately half of the final item pool of 664 items. Thus, each self-report item was tested in 9 cognitive interviews in the first round. After each interview, a staff member reviewed the recording of the interview and the cognitive interviewer's notes from the session singling out the following: (1) items that were understood by everyone and (2) items that were difficult for some participants to answer or which were not interpreted as expected. Overall, by the end of the first round of testing, of the original 664 items, 157 items tested very well, 2 items were omitted, 1 item was split into 2 items, and small modifications were made to many additional items to increase clarity. Sample revised items are presented in Table 2, sample omitted items in Table 3, and sample retained items are presented in Table 4.

For the second round of CI, the 157 items that were understood very clearly were set aside, and 22 participants responded to approximately half of the remaining 506 unique items. Thus, each self-report item in the second round was tested in 11 more cognitive interviews. At the end of round 2, one more item was removed, and minor wording changes were made to several other items.

In the third round of CI, the 157 items that worked well in the first round were added back to the item pool to reassess the entire item pool. In addition, 10 CI sessions were conducted, each on half of the modules as before, so that each item received an additional 5 cognitive interviews. There were virtually no misunderstandings in this third round; less than 1% of items were described as confusing by any participant, and there was only 1 instance in which 2 people misunderstood the same item (this item had a content duplicate and was omitted).

Table 2. Examples of items revised during cognitive interviewing based on: participant think aloud and interviewer probing.

Sample revised items (with *intended diagnostic domain*)	Reason for revision
Original item: I felt anxious. Revised item: I had anxiety. *(Anxiety)*	Participants, particularly those in the South, sometimes defined anxious in the context of "I felt anxious" as excited or eager (eg, "I was anxious to go to the fair"). The noun form, however, did not have the same additional connotation; therefore, the item was revised to use the noun form of anxiety.
Original item: I thought I might be God's personal messenger on Earth. Revised Item: I am the only person who can do God's work on Earth. *(Psychotic disorders-religious delusions)*	The original item produced a high base rate of endorsement among devoutly religious participants. The revised item is distinct from the notion that all people are God's children or messengers.
Original instructions: Now I'm going to ask you about things you thought you might have seen while you were fully awake and it was light . Revised instructions: Now I'm going to ask you about things you might have seen while you were fully awake and there was enough light to see clearly. *(Psychotic disorders-visual hallucinations)*	Participant thinks aloud and interviewer probing responses indicated high endorsement because of the appearance of shadows due to dim light. The revised instructions clarify that visual hallucinations were present when enough light was present to see clearly (ie, eliminate shadows).

Table 3. Examples of items omitted during cognitive interviewing based on: participant think aloud and interviewer probing.

Sample omitted items (with *intended diagnostic domain*)	Reason for omission
I felt the presence of evil around me. *(Psychotic disorders–religious or persecutory delusion)*	Responses from participant think aloud and interviewer probing indicated that participants interpreted the item as meaning there were "bad people" (a bad element) around them, which led to a higher base rate of endorsement than was expected.
People said I did not show emotions. *(Psychotic disorders-affective flattening)*	Participants stated that people did not say this.

Table 4. Examples of items retained during cognitive interviewing based on: participant think aloud and interviewer probing.

Sample retained items (with *intended diagnostic domain*)	Reason for retaining item
I felt sad. *(Depression)*	Easily understood in early rounds of cognitive interviewing.
I had difficulty sitting still. *(Adult ADHD[a])*	Easily understood in early rounds of cognitive interviewing .
I thought I deserved to be punished. *(Psychotic disorders-persecutory delusion or guilt)*	Easily understood in early rounds of cognitive interviewing and central to persecutory delusions.
I felt like my thoughts were being controlled against my will. *(Psychotic disorders-delusions of control)*	Easily understood in early rounds of cognitive interviewing and central to delusions of control.

[a]ADHD: attention deficit hyperactivity disorder.

On conclusion of all 3 rounds of cognitive interviews, the expert panel reviewed the session summaries and agreed on a final set of 661 items that they judged to be clear, concise, and that covered all 13 diagnostic categories. In general, the expert panel erred in keeping items that did well in CI, even if this made for some redundancy as expert panel members knew that the quantitative analysis would enable identification of the most predictive items and allow for future reduction of the item pool.

In the first and second rounds of CI, all 40 participants were also given a clinician-administered SCID. This SCID contained the same modules (and diagnostic categories) that the participants completed in the self-report item pool that included their specific chart diagnosis. The responses to all self-report items were compared with the same participant's responses to the corresponding SCID item(s) to see whether the self-report items would predict the SCID response for the same item or symptom in a real-life application. In all the cases tested, we found that we could identify 1 or more self-report items that predicted each SCID item endorsement. More specifically, where participants selected 4 "often" or 5 "always" on the SAGE-SR (or in negatively scored items, a 1 "never" or 2 "rarely" on the Likert scale), the clinician independently endorsed the associated SCID item on the clinician-administered SCID.

Stage III: Screening Assessment for Guiding Evaluation-Self-Report Instrument Construction and Initial Validation

Eighty-four participants who denied having sought treatment or received medication for a mental illness in the past two years were recruited in Chapel Hill, NC. To recruit participants, study staff passed out flyers describing the study near the campus of a large university and made calls to campus service

organizations to describe the study; some participants were recruited directly by study staff through these efforts and others called in to schedule appointments when they learned about the study secondhand as a result of these recruitment strategies.

The resulting sample ranged in age from 18 to 34 (mean 20.2) years, was 74% female (62/84), 5% African American (4/84), 14% Asian (12/84), 7% Hispanic (6/84), and 68% white (57/84). An additional 5% of participants reported being of more than one race (4/84), and 1 participant declined to provide race information (1% or 1/84). All participants were asked to take the SAGE-SR using a tablet or laptop. A total of 42 participants returned within 7 days (mean 5.24 days) to take the SAGE-SR a second time. The 65-item screener covering 13 domains took an average of 7.3 min to administer to this nonclinical sample, with a standard deviation of 2.4 min. When the follow-up items were taken into consideration, the participants took an average 14 min to take the full SAGE-SR, with a standard deviation of 6.8 min. The Tennessee-based clinical sample was recruited via flyers posted in the clinic waiting room. This sample was comprised of 44 participants who ranged in age from 23 to 76 (mean 47.7) years and were 68% female (30/44). Race data was only available for 66% of this sample (29/44); of those that provided race information, the sample was 69% African American (20/29), 3% Asian (1/29), 14% Hispanic (4/29), 10% white (3/29), and 3% other (1/29). As expected, the screener took participants from the clinical sample longer to complete (average completion time of 9.4 min, with a standard deviation of 3.4 min). The full SAGE-SR took on average 24 min to administer in the public sector clinical sample, with a standard deviation of 12.6 min. In contrast, in research populations, the full NetSCID-CV takes 56 min to administer with a standard deviation of 34 min.

Feedback from the nonclinical sample indicated that participants found the SAGE-SR easy to navigate and complete and found nearly all items clear; one exception was the reference to “unwanted thoughts” in the section on obsessive-compulsive disorder, which participants indicated was too vague and confusing. To increase clarity, a definition was added to the display screen for this item: “Unwanted thoughts are thoughts that kept coming back to you even when you didn't want them to.” The only other feedback regarding clarity was regarding some lead prompts that were intended to prime participants to think of the particular period when they were experiencing the specific symptoms they endorsed during the screener to assess concurrence of the follow-up symptoms with the screener symptoms. For example, the lead prompt for the follow-up questions intended to explore generalized anxiety disorder initially read, “Because of my anxiety or worry,” but participants responded that reverse-scored questions did not work with this phrase; subsequently, the lead prompt phrase was changed to “During the time(s) when I felt anxious…” After this change, the related concurrency items were well understood.

The expert panel convened to review the results from the healthy sample to verify the appropriateness of the screening and diagnostic cutoff criteria. Relatively, few of the nonclinical participants were expected to screen in to take the follow-up questions, and fewer still were expected to meet criteria for inclusion of a diagnosis within the differential. Any items that were endorsed above threshold more than 15% of the time were reviewed by the expert panel. Thresholds for follow-up item administration were intended to be more sensitive, whereas thresholds for diagnosis were intended to be more specific. Minor threshold modifications were made after this review. For example, as mentioned earlier, 3 self-report items represented depressed mood on the screener (“I felt sad,” “I felt depressed,” and “I felt hopeless”); initially, the threshold for screening in to the follow-up depression items was endorsing any of these 3 items as happening at least “sometimes” in the last 30 days. This threshold worked well for the “I felt depressed” and “I felt hopeless” items but was overinclusive for the “I felt sad” item (too many participants screened in), so the screening threshold for that item was changed to at least “often.” In addition, when looking at the consistency with which participants screened in to receive depression follow-up questions, participants who only screened in at 1 time point did so by answering the “I felt sad” screener question at the “sometimes” threshold at that time point; therefore, increasing the threshold for this item also increased the consistency of the screening algorithm.

As part of our preliminary look at quantitative validation, test-retest reliability estimates were calculated for the screening items that were always administered in each of the 13 diagnostic categories covered by the SAGE-SR in the nonclinical sample (the screener section also includes some branching, so all items were not answered by every participant). For the 8 screening modules where the initial screener items included only Likert scale items, we first calculated summary scores within each module and then calculated the intraclass correlation coefficient (ICC) for these summary scores. The screening module for psychotic disorders includes Likert scale items that could indicate hallucinations as well as delusions, so summary scores and ICCs were also calculated for these subcategories of the psychotic disorders screening module. The ICC model used for these analyses was a 2-way mixed model of absolute agreement because the rater was the same at test and retest (self-report). This ICC model was also used to calculate test-retest reliability for the alcohol use disorders screening item, which was a continuous measure of the number of days the participant drank alcohol in the last 30 days. For the remaining 4 modules, categorical items (answered either yes or no) were used for screening purposes (for panic disorder, participants were asked whether they had ever had a panic attack, whereas, for both cannabis and other substance disorders, participants were asked whether they had any use within the past 30 days. For PTSD, participants were asked 4 questions about whether they had (1) ever experienced serious trauma, (2) witnessed serious trauma, (3) had a close friend or relative who was traumatized, or (4) whether they were repeatedly exposed to trauma through their work). For each of these 7 items, we calculated kappa coefficients as a measure of test-retest reliability; however, it was not possible to calculate a kappa coefficient for the diagnostic screening module for other substance use disorders, given that only 1 individual endorsed use in the past 30 days and did so at both time points, leaving empty cells and constants in the 2-way tables. The remaining test-retest reliability results are presented with 2-tailed 95% CIs (using bootstrap methods for the kappa coefficients) in Table 5.

Table 5. Test-retest reliability of 12 diagnostic screening modules of the Screening Assessment for Guiding Evaluation-Self-Report.

Diagnostic screening module	Test-retest reliability	95% CI	*P* value
Depressive disorders[a]	.67	0.46-0.81	<.001
Manic and hypomanic disorders[a]	.50	0.23-0.70	<.001
Generalized anxiety disorder[a]	.60	0.29-0.77	<.001
Panic disorder[b]	.86	0.67-1.00	<.001
Agoraphobia[a,c]	.90	0.82-0.94	<.001
Social anxiety disorder[a]	.83	0.70-0.91	<.001
Obsessive–compulsive disorder[a]	.68	0.33-0.85	<.001
Posttraumatic stress disorder[b]–ever experienced serious trauma	.86	0.63-1.00	<.001
Posttraumatic stress disorder[b]–ever witnessed serious trauma	.60	0.22-0.90	<.001
Posttraumatic stress disorder[b]–close family member or friend experienced serious trauma	.76	0.55-0.95	<.001
Posttraumatic stress disorder[b]–repeated exposure to traumatic events through work	.79	N/A[d,e]	<.001
Adult attention-deficit/hyperactivity disorder[a]	.63	0.27-0.82	<.001
Psychotic disorders [a]	.72	0.41-0.86	<.001
Hallucinations[a]	.65	0.44-0.80	<.001
Delusions[a]	.74	0.31-0.89	<.001
Alcohol use disorder[a]	.70	0.50-0.82	<.001
Cannabis use disorder[b]	.84	0.64-1.00	<.001

[a]Test-retest reliability measure is an intraclass correlation coefficient (2-way mixed model of absolute agreement).

[b]Test-retest reliability measure is a kappa coefficient.

[c]The distribution of summary scores in the agoraphobia domain was highly skewed; a log transformation was performed before calculating the intraclass correlation coefficient for this domain.

[d]N/A: not applicable.

[e]Bootstrap methods were unsuccessful to generate a confidence interval for the kappa coefficient for the posttraumatic stress disorder screening question regarding exposure to trauma through work because of the low base rate of this occurrence in our primarily college student sample.

In determining how to interpret these measures of reliability, we used 2 relevant resources: (1) the presented rationale for interpreting the reliability coefficients used by the researchers conducting the DSM-5 field trials [32,33] and (2) the similar ranges or rationale suggested by Cicchetti [34]. In each of these resources, scores below .60 are considered "fair" or "questionable." Scores from .60 to .75 [34] or .80 [32,33] are considered "good," whereas scores above either .75 or .80 are considered "excellent." Within this framework, test-retest reliabilities for agoraphobia, social anxiety disorder, cannabis use disorder, panic disorder, and 1 (to 3, depending on whether the. 75 or .80 range endpoint is used) of the PTSD items were "excellent," whereas those for depression, GAD, OCD, ADHD, one (to 3) of the PTSD items, psychotic disorders, and the subdomains of hallucinations and delusions were "good." The only domain to not reach at least "good" for test-retest reliability was mania or hypomania, which is consistent with previous attempts to develop self-report items for this diagnostic category [7,8,33].

Discussion

Principal Findings

The SAGE-SR was developed as a self-report alternative to the SCID and NetSCID-CV. The development process included the use of an expert panel to draft and iteratively review items as well as review the results of CI regarding item clarity to ensure that the criteria for 13 diagnostic categories commonly seen in clinical practice were well represented in a final pool of 661 well-understood self-report items. Using this item pool, we constructed the SAGE-SR as a 2-part computerized adaptive assessment with an initial 65-item screening instrument from which respondents who meet screening thresholds branch to follow-up questions to determine which diagnoses are returned for a clinician to consider for differential diagnosis.

Initial validation efforts with a nonclinical sample yielded promising results; qualitative feedback from participants indicated items and instructions were well understood, whereas

the tablet- or laptop-based administration was simple to complete and reasonable in length. Preliminary quantitative validation efforts suggest good consistency in screening algorithms across 2 administration times as well as good to excellent test-retest reliability across all but 1 diagnostic category for the screening items in our small nonclinical sample. The one domain for which test-retest reliability was weakest was mania or hypomania, which has also proven problematic for other researchers attempting to create self-report diagnostic screening assessments [7,8,33]. The expert panel made minor revisions to the mania or hypomania self-report items and screening algorithms; whether these revisions improve the test-retest reliability of these items will be addressed in the results from the ongoing quantitative validation with a larger clinical sample.

Limitations

We believe that the item development and qualitative validation procedures described above were very comprehensive, but although the initial quantitative feedback indicates that the SAGE-SR has great promise, the quantitative results are preliminary and based on a small nonclinical sample. Clearly, the results of this initial validation study will need further replication in a larger clinical sample. Data collection in clinical samples is ongoing, and more extensive quantitative validation will be presented once that work is complete. In addition, as noted earlier, the SCID is typically the gold standard against which the accuracy of most diagnostic assessments is measured. A cross-validation of the SAGE-SR's differential diagnosis against the NetSCID-5-CV's diagnostic algorithms is also currently underway.

Conclusions

The SAGE-SR has an initial diagnostic screener that branches to groups of follow-up items to efficiently produce a differential diagnosis. Because the assessment is self-report, it should be possible to use the SAGE-SR in routine clinical care both in specialty behavioral health and in primary care settings. The SAGE-SR offers the promise of providing a rigorous differential diagnosis based on the SCID-5-CV and DSM-5 to a clinician before their meeting with the client so that their face-to-face time can be focused on clarifying that diagnosis in a manner that builds the rapport so inherent in the success of a therapeutic relationship. Indeed, an additional critique offered against the use of either the SCID-5 or other structured clinical interviews in clinical settings is that, despite the diagnostic rigor they provide, it is difficult to build rapport while adhering to a strict and standardized administration protocol [10].

The SAGE-SR helps address the concerns in the field regarding the need for greater diagnostic rigor as well as assessment of possible comorbidities that might be missed in unstructured clinical interviews while doing so in a cost-effective and clinician time-effective manner. The SAGE-SR also fits into the health care movement exemplified by the personal health record in which patients are empowered to provide information to their clinicians and to participate more actively in determining what treatment is most appropriate for them. The SAGE-SR could help primary care practices satisfy the Affordable Care Act's mandate for screening for depression and alcohol use, while doing so as part of a more comprehensive screen for common behavioral health issues.

In addition to its utility for use in routine clinical care in primary care and specialty behavioral health settings, the SAGE-SR offers rigorous coverage of disorders and utility to clinical researchers as well as for epidemiological studies evaluating large number of participants where clinician-based interviewing is not feasible or is prohibitively expensive. The SAGE-SR covers the same diagnostic categories as the SCID-5-CV and all clinical diagnoses in these categories except for psychiatric diagnoses due to another medical condition and substance-induced diagnoses (see Table 1). Thus, the SAGE-SR covers 28 of the 35 disorders in the 8 primary modules of the SCID-5-CV while taking approximately half as long for respondents to complete and without the training and administration time burdens for the clinician. Like the NetSCID-5-CV, responses to the SAGE-SR populate a detailed database but, unlike the NetSCID-5-CV, the SAGE-SR gathers much more information that could then be available for quantitative analysis. Rather than generating a series of binary criteria endorsements, the SAGE-SR generates a very granular and complete inventory of individual symptoms with Likert scale frequency assessments, thus offering both diagnostic and symptom severity information. This detailed electronic response set can be used to populate admission summaries, progress notes, and discharge summaries, as well as offer a wealth of information on treatment progress and response. The detailed database from the SAGE-SR responses over time can be used to identify the symptom clusters that respond best to specific interventions and maximize the likelihood of measuring change quantitatively to be able to identify best practices.

Given the move toward measurement-based care [35-37], the information provided by the SAGE-SR can potentially be used to look at symptom presentation and severity across multiple time points as well as help clinicians monitor cross-cutting symptoms that might not be part of a primary diagnosis to help justify diagnostic and treatment decisions, fulfilling one of the recommendations of the DSM-5 [33]. In the future, it should also be possible to rescreen clients with the most important items. For example, if OCD, panic disorder, and major depressive episode are included in the differential diagnosis, then the corresponding self-report Likert scale items could be administered at regular intervals. This very focused approach to outcomes tracking should minimize clinician and patient burden. Thus, the SAGE-SR represents a potentially invaluable tool in the move toward measurement-based care.

More information about the SAGE-SR is available on the Web [38] as is a demonstration version of the SAGE-SR [39].

Acknowledgments

This research was supported in part by grants from the National Institutes of Mental Health (RMH095275B, RMH108177A, and R44MH102914) awarded to TeleSage, Inc.

Conflicts of Interest

The NetSCID-5 and the SAGE-SR are sold by TeleSage for profit. BBB, SP, PM, LZ, and MS are employees of TeleSage, Inc. ISB and BBB are the owners of TeleSage. All the other authors declare that they have no conflicts of interest.

References

1. First MB, Williams JB, Karg RS, Spitzer RL. Structured Clinical Interview for DSM-5 Disorders, Clinician Version (SCID-5-CV). Arlington, VA: American Psychiatric Association; 2015.
2. First MB, Williams JB, Karg RS, Spitzer RL. Structured Clinical Interview for DSM-5—Research Version (SCID-5 for DSM-5, Research Version; SCID-5-RV). Arlington, VA: American Psychiatric Association; 2015.
3. Segal DL, Williams K. Structured and semistructured interviews for differential diagnosis: fundamental issues, applications, and features. In: Beidel DC, Frueh BC, Hersen M, editors. Adult Psychopathology and Diagnosis. Hoboken, NJ: John Wiley & Sons Inc; 2014:103-129.
4. Carlbring P, Forslin P, Willebrand M, Ljungstrand P, Strandlund C, Ekselius L, et al. Is the web-administered CIDI-SF equivalent to a human SCID-interview? Eur Psychiatry 2002 May;17:151-152. [doi: 10.1016/S0924-9338(02)80658-7]
5. Germans S, Van Heck GL, Masthoff ED, Trompenaars FJ, Hodiamont PP. Diagnostic efficiency among psychiatric outpatients of a self-report version of a subset of screen items of the Structured Clinical Interview for DSM-IV-TR Personality Disorders (SCID-II). Psychol Assess 2010 Dec;22(4):945-952. [doi: 10.1037/a0021047] [Medline: 21038969]
6. Kessler RC, Abelson J, Demler O, Escobar JI, Gibbon M, Guyer ME, et al. Clinical calibration of DSM-IV diagnoses in the World Mental Health (WMH) version of the World Health Organization (WHO) Composite International Diagnostic Interview (WMHCIDI). Int J Methods Psychiatr Res 2004;13(2):122-139. [Medline: 15297907]
7. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-33;quiz 34. [Medline: 9881538]
8. Zimmerman M, Mattia JI. A self-report scale to help make psychiatric diagnoses: the Psychiatric Diagnostic Screening Questionnaire. Arch Gen Psychiatry 2001 Aug;58(8):787-794. [Medline: 11483146]
9. Lobbestael J, Leurgans M, Arntz A. Inter-rater reliability of the Structured Clinical Interview for DSM-IV Axis I Disorders (SCID I) and Axis II Disorders (SCID II). Clin Psychol Psychother 2011;18(1):75-79. [doi: 10.1002/cpp.693] [Medline: 20309842]
10. Blaine JD, Forman RF, Svikis D. Response: assessing the instruments. Addict Sci Clin Pract 2007 Dec;4(1):31-33 [FREE Full text] [Medline: 18292707]
11. Brodey BB, First M, Linthicum J, Haman K, Sasiela JW, Ayer D. Validation of the NetSCID: an automated web-based adaptive version of the SCID. Compr Psychiatry 2016 Apr;66:67-70 [FREE Full text] [doi: 10.1016/j.comppsych.2015.10.005] [Medline: 26995238]
12. Zimmerman M. What should the standard of care for psychiatric diagnostic evaluations be? J Nerv Ment Dis 2003 May;191(5):281-286. [doi: 10.1097/01.NMD.0000066149.40946.FA] [Medline: 12819546]
13. Bruchmüller K, Margraf J, Suppiger A, Schneider S. Popular or unpopular? Therapists' use of structured interviews and their estimation of patient acceptance. Behav Ther 2011 Dec;42(4):634-643. [doi: 10.1016/j.beth.2011.02.003] [Medline: 22035992]
14. Ramirez Basco M, Bostic JQ, Davies D, Rush AJ, Witte B, Hendrickse W, et al. Methods to improve diagnostic accuracy in a community mental health setting. Am J Psychiatry 2000 Oct;157(10):1599-1605. [doi: 10.1176/appi.ajp.157.10.1599] [Medline: 11007713]
15. Shear MK, Greeno C, Kang J, Ludewig D, Frank E, Swartz HA, et al. Diagnosis of nonpsychotic patients in community clinics. Am J Psychiatry 2000 Apr;157(4):581-587. [doi: 10.1176/appi.ajp.157.4.581] [Medline: 10739417]
16. Zimmerman M, Mattia JI. Psychiatric diagnosis in clinical practice: is comorbidity being missed? Compr Psychiatry 1999;40(3):182-191. [Medline: 10360612]
17. Zimmerman M, Mattia JI. Differences between clinical and research practices in diagnosing borderline personality disorder. Am J Psychiatry 1999 Oct;156(10):1570-1574. [doi: 10.1176/ajp.156.10.1570] [Medline: 10518168]
18. Jensen-Doss A, Weisz JR. Diagnostic agreement predicts treatment process and outcomes in youth mental health clinics. J Consult Clin Psychol 2008 Oct;76(5):711-722. [doi: 10.1037/0022-006X.76.5.711] [Medline: 18837589]
19. Whiteside SP, Sattler AF, Hathaway J, Douglas KV. Use of evidence-based assessment for childhood anxiety disorders in community practice. J Anxiety Disord 2016 Apr;39:65-70 [FREE Full text] [doi: 10.1016/j.janxdis.2016.02.008] [Medline: 26962996]
20. Mechanic D, Olfson M. The relevance of the Affordable Care Act for improving mental health care. Annu Rev Clin Psychol 2016;12:515-542. [doi: 10.1146/annurev-clinpsy-021815-092936] [Medline: 26666969]

21. Widiger TA, Samuel DB. Evidence-based assessment of personality disorders. Psychol Assess 2005 Sep;17(3):278-287. [doi: 10.1037/1040-3590.17.3.278] [Medline: 16262454]
22. World Health Organization. The World Health Organization World Mental Health Composite International Diagnostic Interview (WHO WMH-CIDI) URL: https://www.hcp.med.harvard.edu/wmhcidi/ [accessed 2018-03-06] [WebCite Cache ID 6uwwZHd4M]
23. Lewis G, Pelosi AJ, Araya R, Dunn G. Measuring psychiatric disorder in the community: a standardized assessment for use by lay interviewers. Psychol Med 1992 May;22(2):465-486. [Medline: 1615114]
24. Lewis G. Assessing psychiatric disorder with a human interviewer or a computer. J Epidemiol Community Health 1994 Apr;48(2):207-210 [FREE Full text] [Medline: 8189180]
25. Head J, Stansfeld SA, Ebmeier KP, Geddes JR, Allan CL, Lewis G, et al. Use of self-administered instruments to assess psychiatric disorders in older people: validity of the General Health Questionnaire, the Center for Epidemiologic Studies Depression Scale and the self-completion version of the revised Clinical Interview Schedule. Psychol Med 2013 Dec;43(12):2649-2656 [FREE Full text] [doi: 10.1017/S0033291713000342] [Medline: 23507136]
26. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. J Am Med Assoc 1999 Nov 10;282(18):1737-1744. [Medline: 10568646]
27. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th Edition. Arlington, VA: American Psychiatric Publishing; 2013.
28. Brodey BB, Goodman SH, Baldasaro RE, Brooks-DeWeese A, Wilson ME, Brodey IS, et al. Development of the Perinatal Depression Inventory (PDI)-14 using item response theory: a comparison of the BDI-II, EPDS, PDI, and PHQ-9. Arch Womens Ment Health 2016 Apr;19(2):307-316 [FREE Full text] [doi: 10.1007/s00737-015-0553-9] [Medline: 26271280]
29. Health Measures. Patient-Reported Outcomes Measurement Information System (PROMIS) URL: http://www.healthmeasures.net/explore-measurement-systems/promis [accessed 2018-01-29] [WebCite Cache ID 6wpqM3NEd]
30. Pilkonis PA, Choi SW, Reise SP, Stover AM, Riley WT, Cella D, PROMIS Cooperative Group. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS®): depression, anxiety, and anger. Assessment 2011 Sep;18(3):263-283 [FREE Full text] [doi: 10.1177/1073191111411667] [Medline: 21697139]
31. Willis GB. Cognitive Interviewing: A Tool for Improving Questionnaire Design. Thousand Oaks, CA: Sage Publications; 2005.
32. Kraemer HC, Kupfer DJ, Clarke DE, Narrow WE, Regier DA. DSM-5: how reliable is reliable enough? Am J Psychiatry 2012 Jan;169(1):13-15. [doi: 10.1176/appi.ajp.2011.11010050] [Medline: 22223009]
33. Narrow WE, Clarke DE, Kuramoto SJ, Kraemer HC, Kupfer DJ, Greiner L, et al. DSM-5 field trials in the United States and Canada, Part III: development and reliability testing of a cross-cutting symptom assessment for DSM-5. Am J Psychiatry 2013 Jan;170(1):71-82. [doi: 10.1176/appi.ajp.2012.12071000] [Medline: 23111499]
34. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess 1994;6(4):284-290. [doi: 10.1037/1040-3590.6.4.284]
35. Fortney JC, Unützer J, Wrenn G, Pyne JM, Smith GR, Schoenbaum M, et al. A tipping point for measurement-based care. Psychiatr Serv 2017 Feb 01;68(2):179-188. [doi: 10.1176/appi.ps.201500439] [Medline: 27582237]
36. Lemke S, Boden MT, Kearney LK, Krahn DD, Neuman MJ, Schmidt EM, et al. Measurement-based management of mental health quality and access in VHA: SAIL mental health domain. Psychol Serv 2017 Feb;14(1):1-12. [doi: 10.1037/ser0000097] [Medline: 28134552]
37. Valenstein M, Adler DA, Berlant J, Dixon LB, Dulit RA, Goldman B, et al. Implementing standardized assessments in clinical care: now's the time. Psychiatr Serv 2009 Oct;60(10):1372-1375. [doi: 10.1176/ps.2009.60.10.1372] [Medline: 19797378]
38. Telesage. Screening Assessment for Guiding Evaluation (SAGE-SR) URL: https://telesage.com/sage/ [accessed 2018-03-07] [WebCite Cache ID 6xk3kqMmv]
39. Telesage. Demonstration of the Screening Assessment for Guiding Evaluation (SAGE-SR) URL: https://portal.telesage.com/sage/public/demo [accessed 2018-03-07] [WebCite Cache ID 6xk3vh5qf]

Abbreviations

ADHD: attention deficit hyperactivity disorder
CI: cognitive interviewing
CIDI: Composite International Diagnostic Interview
CV: Clinician Version
DSM: Diagnostic and Statistical Manual of Mental Disorders
GAD: generalized anxiety disorder
ICC: intraclass correlation coefficient
IRB: institutional review board
MINI: Mini International Neuropsychiatric Interview

OCD: obsessive-compulsive disorder
PD: personality disorders
PROMIS: Patient-Reported Outcomes Measurement Information System
PTSD: posttraumatic stress disorder
RV: Research Version
SAGE-SR: Screening Assessment for Guiding Evaluation-Self-Report
SCID: Structured Clinical Interview for DSM
WHO WMH-CIDI: World Health Organization World Mental Health Composite International Diagnostic Interview

Edited by G Eysenbach; submitted 15.11.17; peer-reviewed by TR Soron, J Colditz, T Irizarry; comments to author 16.12.17; revised version received 29.01.18; accepted 13.02.18; published 23.03.18.

Please cite as:
Brodey B, Purcell SE, Rhea K, Maier P, First M, Zweede L, Sinisterra M, Nunn MB, Austin MP, Brodey IS
Rapid and Accurate Behavioral Health Diagnostic Screening: Initial Validation Study of a Web-Based, Self-Report Tool (the SAGE-SR)
J Med Internet Res 2018;20(3):e108
URL: http://www.jmir.org/2018/3/e108/
doi:10.2196/jmir.9428
PMID:29572204

Original Paper

Digital Gaming for Improving the Functioning of People With Traumatic Brain Injury: Randomized Clinical Feasibility Study

Maritta Välimäki[1,2,3], RN, PhD; Kaisa Mishina[1], RN, PhD; Johanna K Kaakinen[4], PhD; Suvi K Holm[4], MA (Psych); Jukka Vahlo[5], MA (Folklore); Markus Kirjonen[4], MA (Psych); Virve Pekurinen[1], RN, MNSc; Olli Tenovuo[6,7], MD, PhD; Jyrki Korkeila[7,8], MD, PhD; Heikki Hämäläinen[4], PhD; Jaana Sarajuuri[9], Lic Psych; Pekka Rantanen[9], MD, PhD; Tage Orenius[10], Lic Psych; Aki Koponen[5], DSc

[1]Department of Nursing Science, Faculty of Medicine, University of Turku, Turku, Finland
[2]Turku University Hospital, Turku, Finland
[3]School of Nursing, Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)
[4]Department of Psychology, Faculty of Social Sciences, University of Turku, Turku, Finland
[5]Turku School of Economics, Centre for Collaborative Research, University of Turku, Turku, Finland
[6]Division of Clinical Neurosciences, Turku University Hospital, Turku, Finland
[7]Faculty of Medicine, University of Turku, Turku, Finland
[8]Psychiatric Care Division, Satakunta Hospital District, Harjavalta, Finland
[9]Validia Rehabilitation Helsinki, Helsinki, Finland
[10]Orton Orthopaedic Hospital, Orton, Helsinki, Finland

Corresponding Author:
Maritta Välimäki, RN, PhD
School of Nursing
Hong Kong Polytechnic University
Hung Hom, Kowloon
Hong Kong,
China (Hong Kong)
Phone: 852 2766 6409
Email: maritta.valimaki@polyu.edu.hk

Abstract

Background: Traumatic brain injury (TBI) is a major health problem that often requires intensive and long-term rehabilitation.

Objective: The aim of this study was to determine whether rehabilitative digital gaming facilitates cognitive functioning and general well-being in people with TBI.

Methods: A total of 90 Finnish-speaking adults with TBI (18-65 years) were recruited from an outpatient neuroscience clinic. The participants were randomly allocated to one of the three groups: a rehabilitation gaming group (n=29, intervention), an entertainment gaming group (n=29, active control), or a passive control group (n=32). The gaming groups were instructed to engage in gaming for a minimum of 30 min per day for 8 weeks. Primary and secondary outcomes were measured at three time points: before the intervention, after the intervention, and 3 months following the intervention. The primary outcome was cognitive status measured by processing speed and visuomotor tasks (The Trail Making Test; Wechsler Adult Intelligence Scale-Fourth Edition, WAIS-IV, symbol search, coding, and cancellation tasks). Secondary outcomes were attention and executive functions (Simon task), working memory (WAIS-IV digit span and Paced Auditory Serial Addition Test, PASAT), depression (Patient Health Questionnaire-9), self-efficacy (General Self-efficacy Scale), and executive functions (Behavior Rating Inventory of Executive Function-Adult Version). Feasibility information was assessed (acceptability, measurement instruments filled, dropouts, adherence, usability, satisfaction, and possible future use). Cognitive measurements were conducted in face-to-face interviews by trained psychologists, and questionnaires were self-administered.

Results: The effects of rehabilitation gaming did not significantly differ from the effects of entertainment gaming or being in a passive control group. For primary outcomes and PASAT tests, the participants in all three groups showed overall improvement in test scores across the three measurement points. However, depression scores increased significantly between baseline and after 8 weeks and between baseline and after 3 months in the rehabilitative gaming group. No differences were found in patients' self-efficacy between the three measuring points in any of the groups. Participants did use the games (rehabilitation group: 93%,

27/29; entertainment group 100%, 29/29). Games were seen as a usable intervention (rehabilitation group: 70%, 14/29; entertainment group: 83%, 20/29). The rehabilitation group was less satisfied with the gaming intervention (68%, 13/29 vs 83%, 20/29), but they were more willing to use the game after the intervention period (76%, 16/29 vs 63%, 15/29). Total time spent on gaming during the intervention period was low (15.22 hour rehabilitation gaming group, 19.22 hour entertainment gaming group).

Conclusions: We did not find differences between the groups in improvement in the outcome measures. The improvements in test performance by all three groups may reflect rehearsal effects. Entertainment gaming had elements that could be considered when rehabilitative games are designed for, implemented in, and assessed in larger clinical trials for persons with TBI.

Trial Registration: ClinicalTrials.gov NCT02425527; https://clinicaltrials.gov/ct2/show/NCT02425527 (Archived by WebCite at http://www.webcitation.org/6esKI1uDH)

(J Med Internet Res 2018;20(3):e77) doi:10.2196/jmir.7618

KEYWORDS

rehabilitation; traumatic brain injury; video games

Introduction

Significance of Traumatic Brain Injury

Traumatic brain injury (TBI) is a major cause of long-term disabilities. In the United States, at least 5.3 million citizens live with disabilities resulting from TBI [1], and in Europe, there is an overall incidence rate of 262 per 100,000 people per year [2]. In Finland (a nation of 5.6 million inhabitants), over 20,000 people suffer from TBI each year [3], and around 100,000 live with disabilities resulting from TBI [4]. Besides emotional symptoms and fatigue [4], cognitive deficits in attention and short-term memory are among the most common and disabling characteristics of people with TBI [5]. The recovery process requires complex, intensive, and long-term–assisted rehabilitation programs [6-8] that inflict a great burden on affected individuals [9] and also on health systems [10]. Individuals with brain injury are suggested to benefit from early and long-term therapeutic interventions [11], and the course of treatment for brain injuries is supported by clinical care guidelines [4,12,13]. Coherent evidence to support the effectiveness of interventions is still scarce [14].

Previous studies have found that rehabilitation interventions after TBI have increased participants' attention, memory, social communication skills, and executive functions [15]. Carney et al [16] concluded in their systematic review, based on two randomized controlled trials (RCTs) and one observational study, that specific forms of cognitive rehabilitation reduce memory failures and anxiety and improve self-concept and interpersonal relationships for persons with TBI. Metacognitive strategy training focusing on functional everyday activities has also been proposed as an appropriate method for rehabilitating people with TBI [17].

Gaming in Improving Functioning of People With Traumatic Brain Injury

A systematic review by Spreij et al [18] has further suggested that computer-based cognitive retraining is one of the most promising novel approaches in improving memory function after an acquired brain injury, although the results are currently inconclusive [19]. Until recently, a majority of patients (75%) with TBI have been younger than 35 years [20]. Likewise, 30% of video game players are in the age range of 18 to 35 years [21], and an average young person has played a total of 10,000 hours of video games by the age of 21 years [22]. It can therefore be assumed that gaming could be a feasible and engaging method in cognitive rehabilitation, especially for young people with TBI. This is supported by the systematic review by Primack et al [23], who concluded that video games have the potential to improve health outcomes in psychological and physical therapy [23]. Gaming has already been used in rehabilitation among people with multiple sclerosis [24], rheumatoid arthritis [25], diabetes [26], complex chronic pain and fatigue [27], spinal cord injury [28], and stroke patients [29,30]. A meta-analytic study of 21 experimental studies by Toril et al [31] indicates that video game training produces positive effects on cognitive functions, including reaction time, attention, memory, and global cognition, although because of the high heterogeneity of the studies, the results must be interpreted with caution. Action video game players have also shown better performance in alertness and cognition compared with those who do not play games [32,33].

Kühn et al [34] found that, for healthy adults, gaming significantly increased gray matter in the right hippocampal formation, right dorsolateral prefrontal cortex, and bilaterally in the cerebellum. The authors concluded that gaming can improve several cognitive functions. Lampit and colleagues [35] reported that computerized cognitive training in elderly healthy adults was modestly effective in improving cognitive performance. However, efficacy varied across cognitive domains and was largely determined by design choices. Bavelier et al [36] have also shown that playing action video games produces significant improvements in attentional control in healthy adults [36]. Furthermore, Ball and colleagues [37] conducted a large-scale cognitive training study and found that, although there was no transfer to other untrained skills, training improved memory, attention, and problem-solving skills. On the other hand, it has been suggested that gaming programs are inadequate for efficient integration in current clinical practice [38,39].

Gaming has already been used to some extent in the rehabilitation of persons with TBI [40]. Vakili et al [41] conducted a controlled study on the effects of video games in the rehabilitation of TBI patients. A total of 31 male TBI patients in the age range of 18 and 65 years were allocated to either a treatment group or a waitlist (treatment-as-usual) control group. The treatment group attended a 2-hour group rehabilitation session once a week for 8 weeks. During these sessions, about

one-quarter of the time was dedicated to psychoeducation and the rest of the time for playing the action video game Medal of Honor: Rising Sun. The treatment group's attentional performance improved in several behavioral measures (namely the Attentional Blink task and some subtasks of the Test of Everyday Attention), as did their self-reported quality of life (QoL; measured with the Comprehensive Quality of Life Scale-Fifth Edition). However, gaming did not have a significant effect on self-reported executive control (as measured by the Behavior Rating Inventory of Executive Functioning-Adult version, BRIEF-A) or self-efficacy (as measured by the General Self-Efficacy Scale, GSE). Although this study with a rather small sample size does showcase gaming as a noteworthy candidate for rehabilitation of TBI, the effect of the psychoeducation part of the treatment is not controlled for.

The scientific evidence for the effectiveness of gaming for enhancing cognitive functioning is mixed at best, and more studies in this area are needed [42,43]. Targeting persons with brain injury is relevant because it is still unknown whether the benefits of video game training can be transferred to clinical settings [34] or whether games can improve cognitive functions important for the management of daily activities [44]. It would also be prudent to explore whether gaming has any positive effects for TBI patients within a broader age range. Previous studies concerning the effects of video gaming on cognition have mostly been conducted with young people [45]. It has also been shown that participants in the age range of 6 and 29 years with acquired brain injury have exhibited significant improvement in processing speed, visual-motor coordination, and response inhibition after playing sessions over 12 weeks with two, 1-hour-long training periods with Nintendo Wii [46]. As TBI often results in long-term disability with adverse social, psychological, and economic consequences, it is important to seek methods that optimize independence and social participation to reduce long-term care needs and enhance QoL [47] for adults with TBI.

In this study, we aim to evaluate the effects and feasibility of digital games for improving cognitive functioning and well-being among people with TBI. We hypothesized that among patients with TBI in the intervention group (rehabilitation gaming), in comparison to the active control group (entertainment gaming) and passive control group, there would be a greater improvement in cognitive functioning (processing speed and visuomotor tasks, attention and executive functions, and working memory) and well-being (depression and self-efficacy).

Methods

Trial Registration

The trial has been registered in trial register ClinicalTrials.gov (NCT02425527).

Design

The study includes a three-arm, parallel, and randomized clinical trial examining the effectiveness and feasibility of digital gaming for improving cognitive functioning and general well-being in people with TBI. The full study design and detailed description of the study methods can be found elsewhere [48].

Sample Size

On the basis of our preliminary power calculations (see [48]), the sample size was expected to be 30 in each group, which is not very strong but reasonable enough for a feasibility study aiming to detect changes within a group between baseline and follow-up outcome measurements with an expected attrition rate close to 0%. However, some patients changed their minds regarding their participation before signing a consent form and dropped out of the recruitment process. Therefore, we needed to recruit more patients (n=106) to have 90 participants for randomization.

Participants and Inclusion and Exclusion Criteria

The study was conducted at the Turku University Hospital, Division of Clinical Neurosciences in Turku, Finland.

The eligibility criteria stipulated that participants must be Finnish-speaking and reading adults, in the age range of 18 and 65 years old, and who have been diagnosed with TBI (ICD-10, S06.X, T90.5). To avoid any confounding factors, they should not have had active participation in cognitive rehabilitation (remediation therapy) during the 3 months before the intervention. To ensure that the participants were comparable regarding their clinical status and able to manage their intervention in their home environment, they should have been discharged from the hospital at least 12 months before the recruitment. In addition, eligible participants had to own a TV and a computer and have Internet access at home.

To see possible effects of the gaming, active digital gamers were excluded, with the allowed gaming time being 5 hours or less per week [36]. Potential participants were also excluded if they had sensory impairment (eg, serious visual impairment), severe cognitive impairment (eg, memory problems, slow processing speed, lack of attention, and linguistic problems), a physical impairment that may restrict the use of computers or computer game control systems unaided (eg, hemiplegia and dysfunction of the central vestibular system), apathy identified in previous neuropsychological evaluations, or a diagnosis of a severe mental disorder (eg, schizophrenia or severe depressive disorders identified as the secondary diagnosis). Patient eligibility was assessed primarily by reviewing Turku University Hospital electronic medical records, after which potentially eligible patients were further interviewed via telephone and assessed face-to-face by a trained psychologist.

Recruitment

The hospital electronic medical records were accessed (June 2015), and patients with a TBI diagnosis were screened to determine which patients fulfilled the eligibility criteria for study participation. Those patients who were assessed to meet the inclusion criteria were contacted by telephone or by mail from June 22, 2015 to November 24, 2015 by researchers. Eligible participants with preliminary interest toward the study received written information about the study by mail in addition to informed consent forms, baseline questionnaires to be filled out, and a short description of the eight entertainment games

(in case of allocation to the entertainment gaming group). They were contacted again after 1 to 2 weeks by telephone to enquire whether they would like to participate in the study and what type of game they would like to play if they were allocated to the group of entertainment gaming. If the contacted individual was interested in participating, the trial manager then received a message (by email, SMS text message [short message service, SMS], or telephone) sent by the recruiting researcher and allocated the patient randomly to one of the three arms of the trial. Eligible participants were then invited to the research laboratory, at which time an informed consent form was signed.

The participants' baseline data were gathered, and cognitive measurements were conducted by a trained psychologist at the test laboratory. At the end of the study, three gaming consoles used in the study were given to randomly chosen study participants.

Randomization and Masking

The study was individually randomized. The randomization and patient allocation were fully centralized (at the University of Turku). An independent trial statistician outside the study group randomly assigned (a block randomization in three blocks) the participants using randomization software (SAS [SAS Institute Inc] for Windows, version 9.3). The randomization list was delivered to the trial manager outside the study group. The trial manager informed the researchers about participants' group after the baseline assessments. The researchers overseeing patient recruitment and randomization were therefore aware of the assignments. Due to the intervention type, allocation was not masked to participants in the intervention and control groups or to researchers who recruited patients. The psychologists, as cognitive outcome assessors, were kept blinded. However, in some occasions, study participants told them about their possible game playing. The data analyst (the trial statistician) was kept blinded to the allocation. As far as we are aware, there was no contact between participants in different groups, as they lived throughout a wide geographic area inside the university hospital catchment area.

Interventions

Rehabilitation Gaming

Patients in the rehabilitation gaming group (intervention group) used an Internet browser–based digital brain training program, CogniFit [49]. We used a Web-based cognitive training platform with 33 games designed with the purpose of improving the user's cognitive abilities as brain exercises. To ensure a user-centered approach, the participants were instructed to play at least one exercise from each of the three categories (memory, spatial perception, and mental planning) during each training session daily, otherwise, they were free to choose which exercises they wished to play. Giving participants a sense of agency aimed to increase the likelihood that participants engaged in gaming as instructed [50].

To support the participants' gaming activities and fidelity for the gaming, written instructions for the rehabilitation game were given to the participant. In addition, instructions for the rehabilitation game were introduced to the participants during the introductory meeting with the researcher (two different researchers, both registered nurses and masters' degree in nursing science), which took about 30 min per person. During the meeting, participants' abilities and previous experience in playing digital games were explored to ensure that the participants had the basic gaming skills required for active gaming. A new email address, a password for the email account, and a personal game account were generated for each participant, as the browser-based program required access through a website, and the user would log in with an email address and a specified password. The participant also tested the game unaided to find out possible barriers in their gaming. To record participants' progression and scores on each of the games, the research team had access to the program, and the progress of each participant's game score was monitored. The participant had also a possibility to monitor their own progress in the program. Information about the frequency of training sessions was also recorded by the participants themselves in a gaming diary. The participants' adherence to and motivation [51] for gaming were supported and monitored by weekly telephone calls. During these telephone calls, participants had also the possibility report any technical problems. The telephone calls were made by two researchers and one research assistant (RA). Researchers had qualifications of registered nurse and masters' degree in nursing science, and the RA had a degree of public health nurse and bachelor's degree in nursing science.

The participants were guided to use the rehabilitation game for at least 30 min per day [34,52,53] over a period of 8 weeks. To encourage, motivate, and hold participants to training, they were supported in planning a schedule for their training sessions (days, time, and frequency) for the entire 8-week gaming period. A more detailed description of the intervention can be found in the study protocol [48].

Entertainment Gaming

Participants in the entertainment gaming group (active control group) used commercial digital games designed for Sony PlayStation 3 (PS3) consoles. The project purchased the participant-selected game (see below) from the official PlayStation Store and downloaded and installed the game into the console given to the participant. Games to be selected by the participants (a total of eight games) were considered to correspond to the rehabilitation games and to contain the same core gameplay elements (see [48]).

As in the intervention group, the participants chose an entertainment game that they found enjoyable, which was assumed to increase the likelihood that participants engage in gaming as instructed and to ensure the attractiveness of the game for the player. However, the participants were not forced to play any one type of game, and they were able to change the game during the 8-week intervention period if they had concerns, for example, because of violent content. Again, during the introductory meeting (about 30 min per person), written instructions regarding how to use the console were given to the participant, and the game the participant selected was tested with the researcher (same researchers as with the rehabilitation gaming group). As with the intervention group, ability to play digital games was explored to ensure that participants had the basic gaming skills required for active gaming. An overview

of the use of the console was also offered, and a tutorial demonstration was given (how to start the console; how to play the game; how to use the controller; how to change game options, such as game difficulty and speed; and so on). A technical assistant was available to visit the participant's home to help set up the console [54] or give guidance by telephone. The participants were guided to play the console for at least 30 min per day over a period of 8 weeks [34]. The participants were supported in planning their training session schedule (days and times), and information about game sessions (day, time, frequency, and play progress) was recorded by the participant in a gaming diary. Therefore, a participant was also able to monitor his or her own progress in the game. Furthermore, adherence to gaming was supported and monitored by weekly telephone calls. As in the intervention group, during these telephone calls, participants had the possibility report any technical problems. The telephone calls were made by the same researchers and RA than in the rehabilitation gaming group.

No Gaming

Patients in the no-gaming group (passive control group) did not have gaming activities organized by the project, but as with the gaming groups, the researchers called them weekly. Participants in this group were offered an opportunity to have games and consoles for a 2-week period free of charge after the study; 11 did so after the follow-up measurement.

Assessment

Patient data were collected at three different times: at baseline, after the intervention (8 weeks, from September 2015 to December 2015), and 3 months after the intervention ended (from December 2015 to April 2016). Cognitive tests were conducted in the research laboratory, and self-administered questionnaires were sent to participants homes to be filled out before each visit to the research laboratory. Participants returned the questionnaires during the visit to the research laboratory.

Outcomes

Primary Outcome

Processing Speed and Visuomotor Tasks

The Trail Making Test (TMT) requires visual search, scanning, speed of processing, mental flexibility, and executive functions [55]. The test consists of two parts, A and B. In TMT A, participants are given a paper displaying circles numbered 1 to 25 in random order; the task is to draw lines that will connect the numbers in ascending order. In TMT B, the circles contain both numbers and letters. In this part of the test, the task is to draw lines to connect the circles so that they alter between numbers and letters in an ascending order (1-A-2-B and so on). The time it takes to complete the trail in each part is recorded (see [48]). TMT and the WAIS-IV subtests are recommended outcome measures in TBI research because of their reliability and validity [56].

In addition to the TMT, three tasks from the WAIS-IV test package were used to further assess processing speed and visuomotor skills: (1) symbol search, (2) cancellation, and (3) coding. The three tasks chosen for this experiment are aimed toward assessing skills of sorting out simple visual information, monitoring, making progress in a task, maintaining attention, visuomotor coordination, and visual memory [57]. In both the symbol search and the cancellation tasks, the participants perform a visual search to find out if a certain symbol is among other symbols. In the symbol search task, the symbols are organized in rows, and the participant must indicate for each row whether or not a required symbol appears on the row. The cancellation task is similar to the symbol search task, but this time the participant seeks to find set symbols during the whole task (instead of the required symbols changing on each row). In the coding task, the participant is given a set of number-symbol pairs. The task is to fill out an empty grid containing only numbers with the appropriate symbols matching those numbers (see [48]).

Secondary Outcomes

Attention and Executive Functions

The Simon task [58,59] was used to measure the inhibition component of executive functions [60]. In the task, a blue or red square appears on either the left or right side of the screen. The participant is instructed to push the left button on a response pad each time a blue square appears and the right button each time a red square appears, irrespective of which side the square is presented. In congruent trials, the response button is on the same side as the square, and in incongruent trials the square is on the opposite side of the response button (ie, the irrelevant spatial information is conflicting with the correct response). The number of correct responses and reaction times are recorded, and the difference between the congruent and incongruent trials is used as a measure of the Simon effect (see [48]).

Working Memory

Working memory was assessed with the digit span task from the WAIS-IV package [57]. In the first part of the task, participants repeat numbers in the order they heard them. In the second part, they repeat the numbers backwards. In the third part of the task, the participants repeat numbers in numerical order [57]. WAIS-IV subtests have been recommended as outcome measures in TBI research because of their reliability and validity [56].

The Paced Auditory Serial Addition Test (PASAT) [61] measures auditory information processing speed, flexibility, and calculation skills [62]. There are two parts in this task. In the first part, single numbers are presented every 3 seconds. The participant adds each new number to the last number before it. In the second part, the numbers are presented every 2 seconds. The test score is the number of correct sums given in each trial (see [48]). PASAT is widely used to assess cognitive changes in TBI patients, and it has good psychometric properties, even though some studies suggest that it may be sensitive to practice effects [63].

Depression

In the Patient Health Questionnaire-9 (PHQ-9) [64,65], a self-administered questionnaire, respondents are asked to indicate how often they have been bothered by any of the problems over the previous 2 weeks, such as little interest or pleasure in doing things, feeling down, depressed, or hopeless,

feeling tired or having little energy. Each of the nine items are scored as 0=not at all, 1=several days, 2=more than half of the days, and 3=nearly every day. On the basis of the individual items, a total score is formed; the higher the score, the more severe the depression symptoms (range: 0-27). The measure has demonstrated diagnostic sensitivity and strong reliability in previous studies among TBI populations [66,67].

Self-Efficacy

The GSE [65] is a self-administered scale that assesses a general sense of perceived self-efficacy to predict coping with daily challenges, as well as adaptation after experiencing a variety of stressful life events. The scale consists of 10 items, and responses are made on a 4-point scale (1=not at all true, 2=hardly true, 3=moderately true, and 4=exactly true). It takes about 4 min to complete. The final composite score ranges from 10 to 40 and comprises the sum of all 10 responses; low scores represent a lower ability to cope with daily problems. The scale has been previously used in studies with TBI populations [68,69].

Executive Functions

The BRIEF-A is a 75-item self-administrated questionnaire that focuses on executive functions in daily life [70]. Responses are given in a 3-point Likert scale (never or sometimes or often), and a global executive composite score is formed by the total score [70] (see [48]). BRIEF-A has been shown to have good psychometric qualities in a sample of TBI patients [71].

Feasibility

Cumulative monitoring was conducted during the 8-week period regarding gaming activities (gaming frequency, timing, and time) in both gaming groups (intervention and active control group). The gaming information concerning the rehabilitation gaming group were collected from game logs retrieved from the gaming system. Regarding entertainment games, the information was collected from console gaming logs where possible and from the gaming diaries where the logs were not available. Feasibility was assessed by collecting the following information during the study process: acceptability as measurement instruments filled out (yes, no; %), attrition as calculating dropouts for any reason (yes, no; %), and adherence as involvement in the interventions for an 8-week period (yes, no; %). Feasibility evaluation in terms of usability, satisfaction, and future use was assessed by asking the participants: Was the game usable? (yes or no; %), Have you been satisfied with the game? (yes or no; %), and Would you like to use the game in the future? (yes or no or maybe; %). The participants had a possibility to specify their answers by answering to open-ended questions (not analyzed in the study because of limited size of the data). In addition, participants' selections of the commercial digital games designed for Sony PS3 consoles are presented.

Background Information

Background information including sociodemographic characteristics and medical history was collected (age, gender, marital status, level of education, employment status, living situation, illness history, and current digital game playing [hours a week]).

Statistical Methods

The sample size needed for the study was based on preliminary estimations (see [48]). The primary and secondary outcomes were assessed at baseline, after the end of the intervention, and 3 months after the intervention ended (at 6 months from baseline). To test the study's hypothesis, the data were analyzed with analysis of covariance, in which Group (rehabilitation gaming, entertainment gaming, and passive control) was a between-subjects factor, and Time (before intervention, after intervention, and 6 months after baseline) was a within-subjects factor. Age was used as a covariate. Effectiveness of the intervention was indicated by a significant Group*Time interaction, which indicated differences between the three groups in the improvement of the primary and secondary outcomes over time.

For sensitivity analysis, we performed analyses for completer-only data and imputer data. We compared the study results between these two groups. No differences between the results were found in the intention to treat analysis or among those completing the follow-ups.

Statistical analyses were performed with SAS system for Windows, version 9.4 and Statistical Package for the Social Sciences (SPSS) statistics version 22 (IBM Corp). *P* values less than .05 are considered statistically significant.

Ethical Issues

The study was evaluated by the Ethics Committee of the Turku University Hospital (ETMK 41/1801/2015), and the permission to conduct the study was granted by Turku University Hospital (T89/T04/008/2015). The trial has been officially registered (NCT02425527). All participants volunteered for the study. The study participants were informed orally (at least two telephone calls and one face-to-face meeting) and in written format of how and where their information was to be accessed, what the purpose of the study was, and what specific steps to be taken were to be (if agreed to participate in the study). Written informed consent was obtained in accordance with the Declaration of Helsinki [72]. To identify any ethical or practical concerns in the study protocol, entertainment and rehabilitation games were pretested with five healthy adults and with five people with TBI. On the basis of pretests, more specific inclusion and exclusion criteria for the study were identified. In addition, some games initially identified to be used in the study were excluded if they were suspected to cause dizziness or headaches because of dark colors or three-dimensional tunnel effects [73].

Results

Sample Characteristics

The flowchart of the participants is described in Figure 1. A total of 758 individuals were screened for eligibility to participate. Of these, 660 were excluded from the study (73 did not meet the inclusion criteria, 203 refused to participate, and 384 could not be contacted by telephone [no answer after two attempts, incorrect or lack of telephone number]). A total of 8 people withdrew their consent: two people refused before randomization, and 6 patients did not show up for the first

research meeting. Out of 106 people recruited, 90 people were randomly allocated to the intervention group (n=29), the active control group (n=29), or the passive control group (n=32). At the follow-up, the response rate for the intervention group was 79%, 86% for the active control group, and 69% for the passive control group (the attrition rate was 21%, 14%, and 31%, respectively).

The mean age of all participants was 41 years, and half (45/90, 50%) were male. Over half of the participants were married (53/90, 59%) or lived with a partner (54/90, 60%). More detailed characteristics of the participants are described in Table 1.

Out of all participants, 43% (39/90) had played digital games weekly before the trial: the highest previous gaming activity was in the rehabilitation gaming group (48%, 14/90), followed by the entertainment gaming group (45%, 13/90), and the passive control group (34%, 11/90). Over half of the participants (57%, 51/90) had not previously played any games (the passive control group: 66%, 21/90; the rehabilitation gaming group: 52%, 15/90; and the entertainment gaming group: 55%, 16/90).

Description of the outcome information at baseline is described in Table 2. Comparisons of the groups at baseline showed no evidence of differences between the groups in any of the measures.

Figure 1. Flowchart.

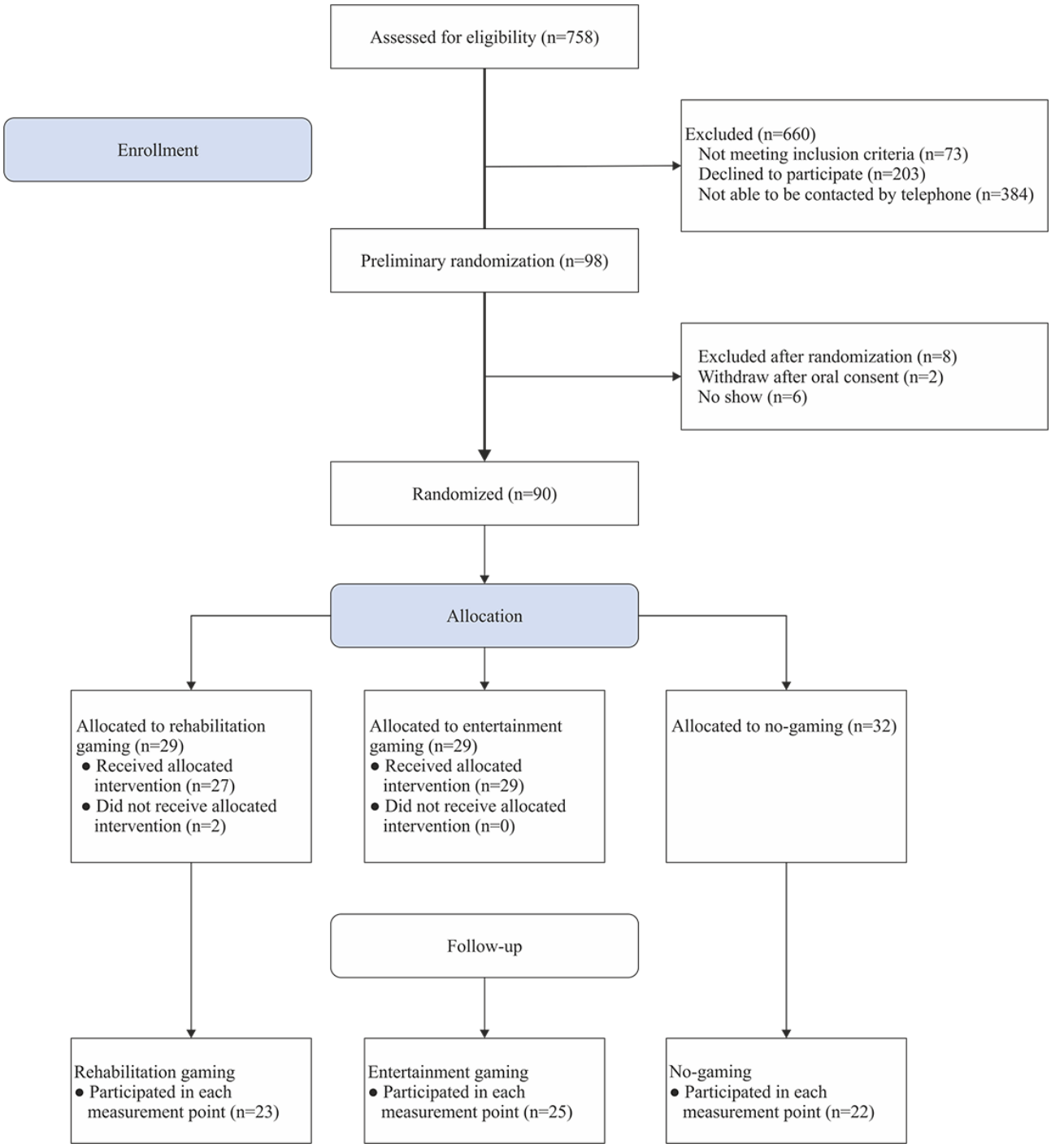


Table 1. Background characteristic of the participants.

Characteristic	Rehabilitation gaming (N=29)	Entertainment gaming (N=29)	No-intervention (N=32)
Age (years), mean (SD)	42.14 (12.15)	40.90 (12.01)	39.34 (12.08)
Sex, n (%)			
Male	15 (52)	13 (45)	17 (53)
Marital status, n (%)			
Single	10 (35)	7 (24)	5 (15)
Married	15 (52)	17 (59)	21 (66)
Divorced	3 (10)	4 (14)	6 (19)
Widowed	0 (0)	1 (3)	0 (0)
Living situation, n (%)			
Alone	10 (35)	11 (38)	9 (28)
With partner	17 (59)	16 (55)	21 (66)
Other	1 (3)	2 (7)	2 (6)
Level of education, n (%)			
No formal education	0 (0)	0 (0)	1 (3)
Secondary grammar school	3 (11)	3 (10)	1 (3)
High school	1 (4)	3 (10)	5 (16)
Vocational education	13 (46)	17 (59)	12 (38)
University degree	11 (40)	6 (21)	12 (37)
Doctoral degree	0 (0)	0 (0)	1 (3)
Employment status, n (%)			
Employed	9 (32)	12 (41)	15 (47)
Retired	10 (36)	10 (35)	8 (25)
Student	3 (11)	1 (3)	5 (16)
Job seeker	2 (7)	2 (7)	1 (3)
Other	4 (14)	4 (14)	3 (9)
Duration of traumatic brain injury (months), mean (SD)	122 (133)	137 (107)	84 (101)

Table 2. Baseline information of different study groups.

Measure	CogniFit		PlayStation 3		Control		*F* (degrees of freedom) value	*P* value
	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)		
TMT[a] A	29	35.28 (15.84)	29	34.31 (12.61)	31	30.45 (12.37)	1.06 (2,86)	.35
TMT B	29	82.83 (54.08)	29	75.72 (24.42)	31	74.87 (28.57)	0.39 (2,86)	.68
WAIS-IV[b] symbol search	29	33.00 (11.32)	29	33.00 (8.21)	31	34.84 (7.71)	0.41 (2,86)	.67
WAIS-IV symbol coding	29	65.00 (21.51)	29	63.79 (12.84)	31	66.84 (16.24)	0.24 (2,86)	.79
WAIS-IV symbol cancellation	29	38.31 (12.40)	29	37.83 (8.06)	31	39.06 (10.39)	0.11 (2,86)	.90
Attention and executive function (Simon test)	29	41.53 (39.94)	28	55.20 (45.64)	31	40.03 (36.76)	1.21 (2,85)	.30
Working memory								
WAIS-IV digital span	29	25.79 (7.07)	29	26.03 (5.23)	31	26.32 (4.69)	0.06 (2,86)	.94
PASAT[c] 3 s	28	43.89 (14.20)	26	44.42 (10.40)	30	46.07 (10.45)	0.27 (2,81)	.77
PASAT 2 s	27	35.59 (10.70)	26	32.38 (8.37)	29	35.48 (9.94)	0.93 (2,79)	.40
Depression (PHQ-9[d])	29	5.79 (3.92)	29	4.66 (5.20)	32	6.34 (6.28)	0.81 (2,87)	.45
Self-efficacy (GSE[e])	29	29.86 (5.55)	29	29.55 (4.77)	32	30.00 (4.87)	0.06 (2,87)	.94
Executive functions (BRIEF-A[f])	20	109.25 (22.02)	20	114.60 (30.75)	21	121.57 (29.40)	1.02 (2,58)	.37

[a]TMT: Trail Making Test.

[b]WAIS-IV: Wechsler Adult Intelligence Scale-Fourth Edition.

[c]PASAT: Paced Auditory Serial Addition Test.

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

[e]GSE: General Self-Efficacy Scale.

[f]BRIEF-A: Behavior Rating Inventory of Executive Functioning-Adult version.

Effects of the Intervention

The analysis of the WAIS-IV *symbol search* scores showed no indication of a Group x Time interaction ($F_{4,138}$=0.34, *P*=.85), which means that we failed to observe differences among groups in test improvement. However, there was a main effect of Time ($F_{2,138}$=4.62, *P*=.01), reflecting overall improvement in test scores across the three measurement points. The main effect age ($F_{1,69}$=5.68, *P*=.02) indicated that older participants had overall lower scores. Furthermore, regarding the WAIS-IV *coding* task, there was no indication of a Group x Time interaction ($F_{4,140}$=1.11, *P*=.35). However, there was a main effect of Time, ($F_{2,140}$=6.29, *P*=.002), indicating overall improvement in the test scores over time. Main effect of age indicated that older participants received lower scores ($F_{1,70}$=6.99, *P*=.01). In the WAIS-IV *cancellation* task scores, there was no evidence for a Group x Time interaction ($F_{4,140}$=0.69, *P*=.60). The only statistically significant effect was the main effect of age ($F_{1,70}$=5.92, *P*=.02), indicating that older participants had lower scores. Thus, although there was overall improvement in the symbol search and coding tasks over time, no differences between groups in the magnitude of the improvement were indicated (see Multimedia Appendix 1).

The results of the TMT version A showed no indication of a two-way interaction between Group and Time ($F_{4,140}$=0.51, *P*=.73), indicating that we failed to observe differences among groups in test improvement in this task. The analysis only indicated a main effect of Age ($F_{1,70}$=11.17, *P*=.001), reflecting overall lower test scores for older participants. In the results of TMT version B, there was no indication of a Group x Time interaction ($F_{3.51,122.77}$=0.31, *P*=.85) or other effects (see Multimedia Appendix 1).

Secondary Outcomes

Attention, Executive Functions and Working Memory

The *Simon task* results revealed no indication of an interaction between Time and Group, ($F_{3.32,114.44}$=0.16, *P*=.94; see Multimedia Appendix 2). There was a main effect of age ($F_{1,69}$=8.76, *P*=.004), indicating that the Simon effect (ie, the difference in the reaction time between congruent and incongruent trials) was greater for older participants.

In the PASAT 3 s version, there was no indication of an interaction between Group and Time ($F_{3.51,108.93}$=1.20, *P*=.33) nor other effects. In the results of the PASAT 2 s version, there was no indication of an interaction ($F_{3.55,110.17}$=0.57, *P*=.67), but there was a main effect of Time ($F_{1.78,110.17}$=9.23, *P*<.001). In other words, we observed overall improvement in the test performance across the three measurement points (see Multimedia Appendix 2).

For the WAIS-IV *digit span* task, there was no indication of an interaction ($F_{4,138}$=0.80, *P*=.53), nor of any main effects (see Multimedia Appendix 2).

Depression

The results for the depression score outcomes are described in Multimedia Appendix 2. The depression score results revealed no indication of an interaction between Time and Group, ($F_{4,134}$=0.848, P=.48). There was no main effect of Time ($F_{2,134}$=2.212, P=.11). The analysis showed that depressive symptoms remained nearly the same between baseline and 3 months, both in the entertainment group (mean 4.12 [SD 4.94] to mean 4.28 [SD 4.67], P=.51) and the passive control group (mean 6.36 [SD 5.91] to mean 6.46 [SD 6.35], P=.84). On the contrary, in the rehabilitation gaming group, the mean scores increased from baseline to 8 weeks and from baseline to 3 months (mean 5.04 [SD 3.82] to mean 6.65 [SD 5.00], P=.05), showing increase in the participants' depressive symptoms. On the categorical level (mild vs moderate depression), the change observed between time points was not clinically significant (scoring 6-9 points indicates minimal symptoms, University of Michigan Health System (UMHS) Depression Guideline, August 2011).

Self-Efficacy

The general self-efficacy scores results revealed no indication of an interaction between Time and Group, ($F_{3.77,126.19}$=0.534, P=.70). There was no main effect of Time ($F_{1.88,129.19}$=1.38, P=.26). Self-efficacy among the participants increased slightly in the passive control group (P=.06) over time, but the change was not statistically significant (see Multimedia Appendix 2).

Executive Functions

The analysis of the BRIEF-A scores revealed no indication of an interaction between Time and Group, ($F_{4,114}$=1.99, P=.10). The main effect of Time ($F_{2,114}$=3.54, P=.03) indicated that there were differences between measurement points (see Multimedia Appendix 2). A post hoc comparison between baseline and 8 weeks after the intervention was conducted. No statistically significant differences between these two time points were found.

Feasibility

The most favorable game was Ratchet and Clank—Tools of Destruction (26/90) and the Last of us (14/90). The least favorable games were Beyond Good and Evil and Batman: Arkham City (Figure 2).

Out of 758 patients screened, the refusal rate was 27%. A total of 20 randomized participants dropped out of the study during the intervention period (attrition rate 21% (6/29) in the rehabilitation group, 14% (4/29) in the entertainment group, and 31% (10/32) in the passive control group; Table 3).

During the 8-week intervention period, the average gaming time in the entertainment gaming group was 19.22 hours (range 0-71.48 hours) and in the rehabilitation gaming group 15.02 hours (range 0.12-71.38 hours). In general, the participants were adherent to the intervention (entertainment group 100%, 29/29, rehabilitation group 93%, 27/29), and they attended the prescheduled testing sessions (86%, 25/29, in the entertainment group, 79%, 23/29, in the rehabilitation group). Most participants in the entertainment group (83%, 20/29) and in the rehabilitation group (70%, 14/29) also agreed that the usability of the gaming was good, and about two-thirds (68%, 13/29) of the rehabilitation group and 83% (20/29) of the entertainment group were satisfied with the game. Contrary to our expectations, more participants in the rehabilitation group than the entertainment group were willing to use the type of game they were assigned after the intervention was finished as part of their rehabilitation process (76%, 16/29, vs 63%, 15/29; Figure 3).

Figure 2. Selections of entertainment games by participants (each participant could change their game and therefore select more than one game).

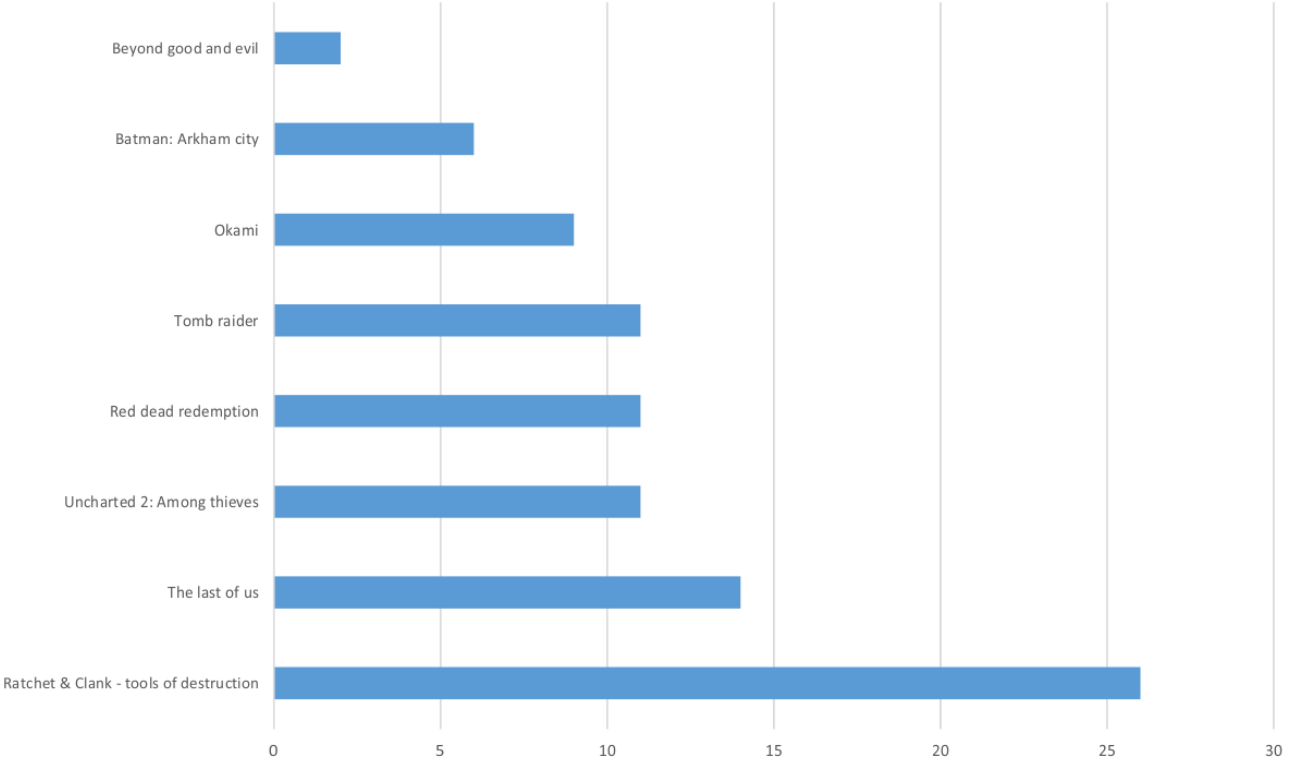
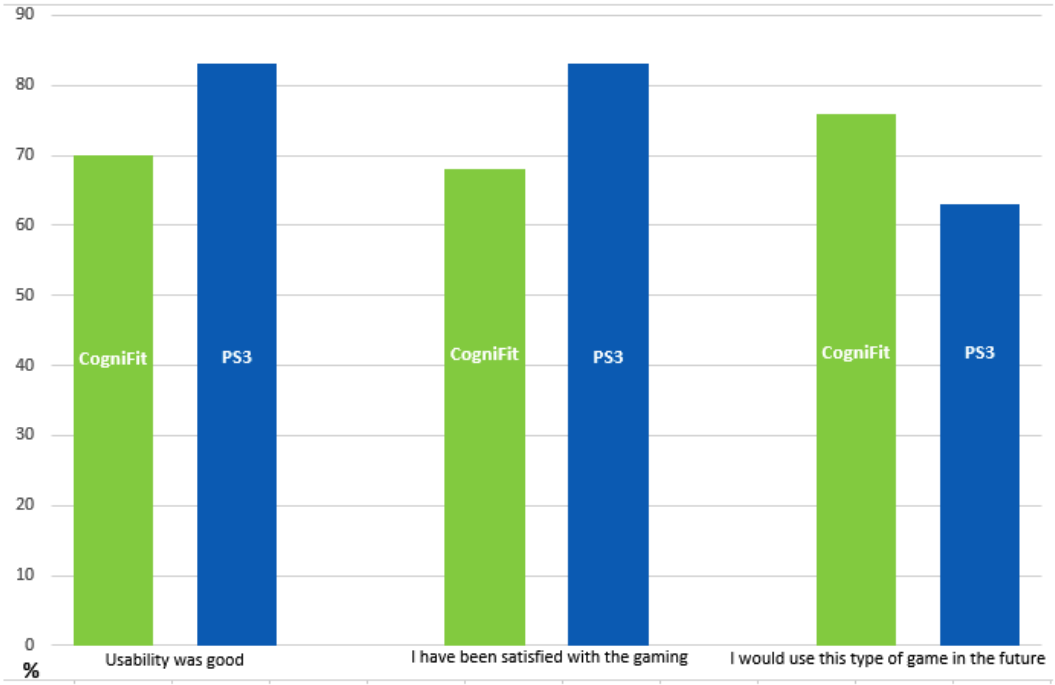


Table 3. Criteria for the feasibility of rehabilitation and entertainment gaming.

Criterion	Rehabilitation gaming group (n=29), n (%)	Entertainment gaming group (n=29), n (%)
Adherence		
Prescheduled measurements performed	23 (79)	25 (86)
Participants' dropout because of any reason	6 (21)	4 (14)
The acceptability of the game	28 (95)	(29) 100
Usability		
Usability evaluation for the gaming system (<80%)	14 (70)	20 (83)
Satisfaction		
Satisfied with the games (<80%)	13 (68)	20 (83)
Use in the future		
Willing to use the games later as part of their recovery process (<60%)	16 (76)	15 (63)

Figure 3. Comparison of the usability, satisfaction, and future use between rehabilitation gaming and entertainment gaming groups.



Discussion

Principal Findings

We evaluate the effects and feasibility of digital games for improving cognitive functioning and well-being among people with TBI. We found no differences between the control group and the two intervention groups for the primary outcomes (processing speed and visuomotor tasks) or any of the secondary outcomes. Test scores improved in all groups over time regarding several different variables.

As this improvement is not related to gaming (ie, whether the participants belonged to one of the two gaming groups or the control group), this effect was likely due to practice effects on the tasks used for assessing the outcomes, which compromises attempts to detect between-group differences in the improvement over time [74]. Indeed, the improvement in the postgaming performance was limited to the cognitive tasks and was not observed in, for example, depression symptoms. However, it is worth noting that in the rehabilitation group, the respondents' depressive symptoms increased during the study period. This finding is important when a new intervention is introduced to the participants. We must, therefore, question whether gaming is associated with this increase of depressive symptoms, even though there is evidence of an association between depression symptoms or mood disorders and TBI [75-77]. Thus, in the future, studies should focus on how gaming might affect persons with TBI, to avoid any harm in patients' clinical status and their QoL.

The age range in our study participants was broad (18-65 years), and our sample size was relatively small. Our study results would have been different if we would have narrowed the age of participants and aimed for a younger target population. This could be reasonable adjustment, keeping in mind that the

majority of patients (75%) with TBI are younger than 35 years [20]. A systematic review of Shams et al [45] also showed that most studies aiming to improve cognitive functioning have targeted younger participants. On the other hand, we aimed to capture real-life events of individuals from a variety of age groups, who were not active gamers. We also wanted to facilitate gaming at home to increase participants' engagement in intervention [78]. However, this decision may have caused another concern. We found low fidelity in the intervention, which may be a result of participants' independent gaming intervention at home. Patients in the study by Lampit and colleagues [35] also found that training in an unsupervised home environment is not as effective as supervised training. On the other hand, the differences between the rehabilitation and entertainment game groups in adherence to the intervention and the experienced usability and satisfaction with the games could in part be related to differences between the samples: the rehabilitation group subjects were slightly more often single and less often employed than entertainment game group subjects. These factors may be related to depression, which also slightly increased during the study period in the rehabilitation game group. We observed a high attrition rate, especially in the passive control group, that is, 31%. This has also been found in earlier studies related to cognitive rehabilitation and TBI, such as the study by Vakili & Langdon (2016) with high attrition rate among control group [41]. Dropout in previous studies has been caused by fatigue [79], mental fatigue, or headache during computerized rehabilitation [80]. In our study, the reasons for dropout have not been systematically collected. On the basis of previous studies, to avoid loss of motivation in long training schedules, shorter gaming interventions (1-6 weeks) might be more beneficial for older adults [31]. The training may be seen as exciting at first, but may later be considered boring [81]. Furthermore, we are unaware of how many participants in the control group were engaged in gaming activities, a factor that could positively affect their cognitive status. Of course, there is the possibility that gaming simply is not effective in the rehabilitation of TBI. In addition, because of system updates in the intervention group, some technical errors appeared, and some functions in participants' rehabilitation game user accounts changed (eg, prespecified gaming categories in CogniFit were not included in the participants' user accounts). These changes affected 13 participants for 11 days, and it is therefore unlikely that this affected the results. Furthermore, even though we did include sensitive reaction time measurements among our secondary outcomes (ie, the Simon task), our primary measures were paper-and-pencil tests, which may not have been sensitive enough to capture subtle improvements in cognitive performance. In addition, the association of self-monitoring data concerning the participants' gaming progress and the effects of the gaming could also be used to assess cofounding factors.

In the entertainment gaming group, the participants were given the chance to select their favorite game or change the game during the intervention. Although all the games included in this study were considered to contain similar game dynamics assumed to improve certain cognitive functions, there was some variability between the games, and it is possible that the game dynamics of the participant's favorite game did not target the specific cognitive deficits of that participant. The choice of eight games also makes it difficult to conclude which types of game dynamics actually improve the cognitive functions of interest. In the future, a single game might be a better option in RCT design to ensure accuracy of the content of the different interventions.

The games used in this study may have also included too many action games, whereas participants might prefer other types of games. Vahlo et al [82] showed in their study that only about one-fifth of healthy adults enjoyed playing action-adventure games. As the dislike toward certain types of games or game activities can be rather strong, game selection should represent a wider variety of different game genres to meet personalized gaming preferences in future trials. Some incidental factors could also have shown to affect the outcomes of the results, such as patient perceptions or attitudes toward gaming. As far as we are aware, technological solutions are not used routinely in outpatient care for persons with TBI. If the participants do not see gaming as a seriously taken opportunity for rehabilitation, its effects may be questioned, and recommended gaming instructions may not be followed. We did not perform expectancy testing before the assigned intervention, which could be useful to avoid a placebo effect. We are not aware if patients' prior expectancies of the effectiveness of the gaming affected the outcome of the study [83]. Therefore, the participants' own perceptions toward gaming and its use as part of rehabilitative interventions should be explored in more detail. In the future, patients could potentially be prescribed personalized gaming interventions based on specific cognitive deficits and their personal game preferences, which would improve the effectiveness of the intervention.

Finally, the sample size of the study was small, making it difficult to detect small effects (ie, differences between groups), especially as the sample included a relatively heterogeneous group of patients with a wide variety of cognitive deficits. These factors limit the generalization of the results to a wider population. In future studies, a research design with a larger sample size is needed.

Conclusions

To receive valid outcomes of the effectiveness of gaming, it would be important to make sure that the gaming dose is high enough. One way to do this is to ensure participants' gaming is monitored daily.

Acknowledgments

The authors would like to thank Leigh Ann Lindholm, MA, who provided them with language assistance. They would also like to thank Jussi Välimaki in his effort to offer technical support for the participants. They would also like to show their appreciation to the partners who support the project, including The Finnish Funding Agency for Technology and Innovation, The Finnish

Association of People with Physical Disabilities, Oy Nordisk Film Ab/Playstation, TribeFlame Oy, BCB Medical, Validia Rehabilitation Helsinki, State Research Funding (EVO) assigned by the Ministry of Social Affairs and Health, ORTON Ltd, ORTON Foundation, Serious Games Finland Oy, the Hospital District of Southwest Finland, and Turku University Hospital. The parties responsible for the funding of this study had no role in the study design, data collection, data analysis, data interpretation, or writing of the manuscript

Conflicts of Interest

None declared.

Multimedia Appendix 1

Analysis of primary outcome, processing speed, and visuomotor task, on baseline, 8 weeks, and 3 months (analysis of covariance).

[PDF File (Adobe PDF File), 32KB - jmir_v20i3e77_app1.pdf]

Multimedia Appendix 2

Analysis of secondary outcomes at baseline, 8 weeks, and 3 months (analysis of covariance).

[PDF File (Adobe PDF File), 39KB - jmir_v20i3e77_app2.pdf]

Multimedia Appendix 3

CONSORT‐EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 100KB - jmir_v20i3e77_app3.pdf]

References

1. Langlois JA, Rutland-Brown W, Wald MM. The epidemiology and impact of traumatic brain injury: a brief overview. J Head Trauma Rehabil 2006;21(5):375-378. [Medline: 16983222]
2. Peeters W, van den BR, Polinder S, Brazinova A, Steyerberg EW, Lingsma HF, et al. Epidemiology of traumatic brain injury in Europe. Acta Neurochir (Wien) 2015 Oct;157(10):1683-1696 [FREE Full text] [doi: 10.1007/s00701-015-2512-7] [Medline: 26269030]
3. Gustavsson A, Svensson M, Jacobi F, Allgulander C, Alonso J, Beghi E, CDBE2010Study Group. Cost of disorders of the brain in Europe 2010. Eur Neuropsychopharmacol 2011 Oct;21(10):718-779. [doi: 10.1016/j.euroneuro.2011.08.008] [Medline: 21924589]
4. Kaypahoito.fi. 2008. [Brain injuries: current care guideline] URL: http://www.kaypahoito.fi/web/kh/suositukset/suositus?id=hoi18020 [accessed 2017-11-15] [WebCite Cache ID 6oAfqWRP5]
5. Slovarp L, Azuma T, Lapointe L. The effect of traumatic brain injury on sustained attention and working memory. Brain Inj 2012;26(1):48-57. [doi: 10.3109/02699052.2011.635355] [Medline: 22149444]
6. Van Peppen RP, Kwakkel G, Wood-Dauphinee S, Hendriks HJ, Van der Wees PJ, Dekker J. The impact of physical therapy on functional outcomes after stroke: what's the evidence? Clin Rehabil 2004 Dec;18(8):833-862. [Medline: 15609840]
7. Cramer SC. Repairing the human brain after stroke: I. Mechanisms of spontaneous recovery. Ann Neurol 2008 Mar;63(3):272-287. [doi: 10.1002/ana.21393] [Medline: 18383072]
8. Huda S, Rodriguez R, Lastra L, Warren M, Lacourse MG, Cohen MJ, et al. Cortical activation during foot movements: II effect of movement rate and side. Neuroreport 2008 Oct 29;19(16):1573-1577. [doi: 10.1097/WNR.0b013e328311ca1c] [Medline: 18845938]
9. LoPresti EF, Bodine C, Lewis C. Assistive technology for cognition. IEEE Eng Med Biol Mag 2008;27(2):29-39. [doi: 10.1109/EMB.2007.907396] [Medline: 18463019]
10. World Health Organization (WHO). 2015. Violence and Injury Prevention and Disability (VIP): Neurotrauma URL: http://www.who.int/violence_injury_prevention/road_traffic/activities/neurotrauma/en/ [WebCite Cache ID 6oAgofv7u]
11. Mackay LE, Bernstein BA, Chapman PE, Morgan AS, Milazzo LS. Early intervention in severe head injury: long-term benefits of a formalized program. Arch Phys Med Rehabil 1992 Jul;73(7):635-641. [Medline: 1622318]
12. Marshall S, Bayley M, McCullagh S, Velikonja D, Berrigan L. Clinical practice guidelines for mild traumatic brain injury and persistent symptoms. Can Fam Physician 2012 Mar;58(3):257-267 [FREE Full text] [Medline: 22518895]
13. Scottish Intercollegiate Guidelines Network (SIGN). Edinburgh; 2013. Brain injury rehabilitation in adults URL: http://www.sign.ac.uk/sign-130-brain-injury-rehabilitation-in-adults.html [WebCite Cache ID 6oAimtqt7]
14. High WM, Roebuck-Spencer T, Sander AM, Struchen MA, Sherer M. Early versus later admission to postacute rehabilitation: impact on functional outcome after traumatic brain injury. Arch Phys Med Rehabil 2006 Mar;87(3):334-342. [doi: 10.1016/j.apmr.2005.11.028] [Medline: 16500166]

15. Cicerone KD, Langenbahn DM, Braden C, Malec JF, Kalmar K, Fraas M, et al. Evidence-based cognitive rehabilitation: updated review of the literature from 2003 through 2008. Arch Phys Med Rehabil 2011 Apr;92(4):519-530. [doi: 10.1016/j.apmr.2010.11.015] [Medline: 21440699]
16. Carney N, Chesnut RM, Maynard H, Mann NC, Patterson P, Helfand M. Effect of cognitive rehabilitation on outcomes for persons with traumatic brain injury: a systematic review. J Head Trauma Rehabil 1999 Jun;14(3):277-307. [Medline: 10381980]
17. Ponsford J, Bayley M, Wiseman-Hakes C, Togher L, Velikonja D, McIntyre A, et al. INCOG recommendations for management of cognition following traumatic brain injury, part II: attention and information processing speed. J Head Trauma Rehabil 2014;29(4):321-337. [doi: 10.1097/HTR.0000000000000072] [Medline: 24984095]
18. Spreij LA, Visser-Meily JM, van Heugten CM, Nijboer TC. Novel insights into the rehabilitation of memory post acquired brain injury: a systematic review. Front Hum Neurosci 2014;8:993 [FREE Full text] [doi: 10.3389/fnhum.2014.00993] [Medline: 25566021]
19. Politis AM, Norman RS. Computer-based cognitive rehabilitation for individuals with traumatic brain injury: a systematic review. Perspect ASHA Spec Interest Groups 2016 Mar 31;1(2):18-46. [doi: 10.1044/persp1.SIG2.18]
20. Tagliaferri F, Compagnone C, Korsic M, Servadei F, Kraus J. A systematic review of brain injury epidemiology in Europe. Acta Neurochir (Wien) 2006 Mar;148(3):255-268. [doi: 10.1007/s00701-005-0651-y] [Medline: 16311842]
21. Statista. 2016. Age breakdown of video game players in the United States in 2016 URL: https://www.statista.com/statistics/189582/age-of-us-video-game-players-since-2010/ [accessed 2017-02-10] [WebCite Cache ID 6oAj8Q77P]
22. McGonigal J. Reality Is Broken. 1st edition. New York: Penguin Books; 2011.
23. Primack BA, Carroll MV, McNamara M, Klem ML, King B, Rich M, et al. Role of video games in improving health-related outcomes: a systematic review. Am J Prev Med 2012 Jun;42(6):630-638 [FREE Full text] [doi: 10.1016/j.amepre.2012.02.023] [Medline: 22608382]
24. Taylor MJ, Griffin M. The use of gaming technology for rehabilitation in people with multiple sclerosis. Mult Scler 2015 Apr;21(4):355-371. [doi: 10.1177/1352458514563593] [Medline: 25533296]
25. Allam A, Kostova Z, Nakamoto K, Schulz PJ. The effect of social support features and gamification on a Web-based intervention for rheumatoid arthritis patients: randomized controlled trial. J Med Internet Res 2015;17(1):e14 [FREE Full text] [doi: 10.2196/jmir.3510] [Medline: 25574939]
26. Lee S, Shin S. Effectiveness of virtual reality using video gaming technology in elderly adults with diabetes mellitus. Diabetes Technol Ther 2013 Jun;15(6):489-496. [doi: 10.1089/dia.2013.0050] [Medline: 23560480]
27. Vugts MA, Joosen MC, van Bergen AH, Vrijhoef HJ. Feasibility of applied gaming during interdisciplinary rehabilitation for patients with complex chronic pain and fatigue complaints: a mixed-methods study. JMIR Serious Games 2016 Apr 01;4(1):e2 [FREE Full text] [doi: 10.2196/games.5088] [Medline: 27036094]
28. Betker AL, Desai A, Nett C, Kapadia N, Szturm T. Game-based exercises for dynamic short-sitting balance rehabilitation of people with chronic spinal cord and traumatic brain injuries. Phys Ther 2007 Oct;87(10):1389-1398 [FREE Full text] [doi: 10.2522/ptj.20060229] [Medline: 17712036]
29. Cheok G, Tan D, Low A, Hewitt J. Is nintendo wii an effective intervention for individuals with stroke? A systematic review and meta-analysis. J Am Med Dir Assoc 2015 Nov 01;16(11):923-932. [doi: 10.1016/j.jamda.2015.06.010] [Medline: 26253322]
30. Paquin K, Ali S, Carr K, Crawley J, McGowan C, Horton S. Effectiveness of commercial video gaming on fine motor control in chronic stroke within community-level rehabilitation. Disabil Rehabil 2015;37(23):2184-2191. [doi: 10.3109/09638288.2014.1002574] [Medline: 25586794]
31. Toril P, Reales JM, Ballesteros S. Video game training enhances cognition of older adults: a meta-analytic study. Psychol Aging 2014 Sep;29(3):706-716. [doi: 10.1037/a0037507] [Medline: 25244488]
32. Green CS, Bavelier D. Action video game modifies visual selective attention. Nature 2003 May 29;423(6939):534-537. [doi: 10.1038/nature01647] [Medline: 12774121]
33. Green CS, Bavelier D. Enumeration versus multiple object tracking: the case of action video game players. Cognition 2006 Aug;101(1):217-245 [FREE Full text] [doi: 10.1016/j.cognition.2005.10.004] [Medline: 16359652]
34. Kühn S, Gleich T, Lorenz RC, Lindenberger U, Gallinat J. Playing Super Mario induces structural brain plasticity: gray matter changes resulting from training with a commercial video game. Mol Psychiatry 2014 Feb;19(2):265-271. [doi: 10.1038/mp.2013.120] [Medline: 24166407]
35. Lampit A, Hallock H, Valenzuela M. Computerized cognitive training in cognitively healthy older adults: a systematic review and meta-analysis of effect modifiers. PLoS Med 2014 Nov;11(11):e1001756 [FREE Full text] [doi: 10.1371/journal.pmed.1001756] [Medline: 25405755]
36. Bavelier D, Green CS, Pouget A, Schrater P. Brain plasticity through the life span: learning to learn and action video games. Annu Rev Neurosci 2012;35:391-416. [doi: 10.1146/annurev-neuro-060909-152832] [Medline: 22715883]
37. Ball K, Berch DB, Helmers KF, Jobe JB, Leveck MD, Marsiske M, Advanced Cognitive Training for Independent and Vital Elderly Study Group. Effects of cognitive training interventions with older adults: a randomized controlled trial. J Am Med Assoc 2002 Nov 13;288(18):2271-2281 [FREE Full text] [Medline: 12425704]

38. Schoenberg MR, Ruwe WD, Dawson K, McDonald NB, Houston B, Forducey PG. Comparison of functional outcomes and treatment cost between a computer-based cognitive rehabilitation teletherapy program and a face-to-face rehabilitation program. Prof Psychol Res Pr 2008;39(2):169-175. [doi: 10.1037/0735-7028.39.2.169]
39. Kueider AM, Parisi JM, Gross AL, Rebok GW. Computerized cognitive training with older adults: a systematic review. PLoS One 2012;7(7):e40588 [FREE Full text] [doi: 10.1371/journal.pone.0040588] [Medline: 22792378]
40. Shapi'i A, Mat Zin NA, Elaklouk AM. A game system for cognitive rehabilitation. Biomed Res Int 2015;2015:493562 [FREE Full text] [doi: 10.1155/2015/493562] [Medline: 25815320]
41. Vakili A, Langdon R. Cognitive rehabilitation of attention deficits in traumatic brain injury using action video games: a controlled trial. Cogent Psychol 2016 Feb 24;3(1):493562 [FREE Full text] [doi: 10.1080/23311908.2016.1143732]
42. Gravel J, D'Angelo A, Carrière B, Crevier L, Beauchamp MH, Chauny J, et al. Interventions provided in the acute phase for mild traumatic brain injury: a systematic review. Syst Rev 2013;2:63 [FREE Full text] [doi: 10.1186/2046-4053-2-63] [Medline: 23924958]
43. Lohse KR, Hilderman CG, Cheung KL, Tatla S, Van der Loos HF. Virtual reality therapy for adults post-stroke: a systematic review and meta-analysis exploring virtual environments and commercial games in therapy. PLoS One 2014;9(3):e93318 [FREE Full text] [doi: 10.1371/journal.pone.0093318] [Medline: 24681826]
44. Boot WR, Kramer AF, Simons DJ, Fabiani M, Gratton G. The effects of video game playing on attention, memory, and executive control. Acta Psychol (Amst) 2008 Nov;129(3):387-398. [doi: 10.1016/j.actpsy.2008.09.005] [Medline: 18929349]
45. Shams TA, Foussias G, Zawadzki JA, Marshe VS, Siddiqui I, Müller DJ, et al. The effects of video games on cognition and brain structure: potential implications for neuropsychiatric disorders. Curr Psychiatry Rep 2015 Sep;17(9):71. [doi: 10.1007/s11920-015-0609-6] [Medline: 26216589]
46. de Kloet AJ, Berger MA, Verhoeven IM, van Stein CK, Vlieland TP. Gaming supports youth with acquired brain injury? A pilot study. Brain Inj 2012;26(7-8):1021-1029. [doi: 10.3109/02699052.2012.654592] [Medline: 22632604]
47. von Steinbuechel N, Wilson L, Gibbons H, Muehlan H, Schmidt H, Schmidt S, et al. QOLIBRI overall scale: a brief index of health-related quality of life after traumatic brain injury. J Neurol Neurosurg Psychiatry 2012 Nov;83(11):1041-1047. [doi: 10.1136/jnnp-2012-302361] [Medline: 22851609]
48. Välimäki M, Korkeila J, Kauppi K, Kaakinen JK, Holm S, Vahlo J, et al. Digital gaming for improving the functioning of people with traumatic brain injury: protocol of a feasibility study. JMIR Res Protoc 2016 Feb 09;5(1):e6 [FREE Full text] [doi: 10.2196/resprot.4841] [Medline: 26860741]
49. Cognifit. 2017. Brain Training, Brain Games and Brain Fitness URL: https://www.cognifit.com/ [accessed 2017-02-10] [WebCite Cache ID 6oAjKzoo2]
50. Whitton N. Digital Games and Learning: Research and Theory. 1st edition. New York & London: Routledge; 2014.
51. Ryan RM, Rigby CS, Przybylski A. The motivational pull of video games: a self-determination theory approach. Motiv Emot 2006 Nov 29;30(4):344-360. [doi: 10.1007/s11031-006-9051-8]
52. Li R, Polat U, Makous W, Bavelier D. Enhancing the contrast sensitivity function through action video game training. Nat Neurosci 2009 May;12(5):549-551 [FREE Full text] [doi: 10.1038/nn.2296] [Medline: 19330003]
53. Li R, Polat U, Scalzo F, Bavelier D. Reducing backward masking through action game training. J Vis 2010;10(14):1-13. [doi: 10.1167/10.14.33] [Medline: 21191129]
54. Anguera JA, Boccanfuso J, Rintoul JL, Al-Hashimi O, Faraji F, Janowich J, et al. Video game training enhances cognitive control in older adults. Nature 2013 Sep 5;501(7465):97-101 [FREE Full text] [doi: 10.1038/nature12486] [Medline: 24005416]
55. Tombaugh TN. Trail Making Test A and B: normative data stratified by age and education. Arch Clin Neuropsychol 2004 Mar;19(2):203-214. [doi: 10.1016/S0887-6177(03)00039-8] [Medline: 15010086]
56. Wilde EA, Whiteneck GG, Bogner J, Bushnik T, Cifu DX, Dikmen S, et al. Recommendations for the use of common outcome measures in traumatic brain injury research. Arch Phys Med Rehabil 2010 Nov;91(11):1650-1660.e17. [doi: 10.1016/j.apmr.2010.06.033] [Medline: 21044708]
57. Wechsler D. Pearsonclinical. 2008. Wechsler Adult Intelligence Scale (WAIS-IV). 4th edition URL: https://www.pearsonclinical.com/psychology/products/100000392/wechsler-adult-intelligence-scalefourth-edition-wais-iv.html [accessed 2018-02-15] [WebCite Cache ID 6xFqH3VlG]
58. Simon JR, Rudell AP. Auditory S-R compatibility: the effect of an irrelevant cue on information processing. J Appl Psychol 1967 Jun;51(3):300-304. [Medline: 6045637]
59. Soveri A, Rodriguez-Fornells A, Laine M. Is there a relationship between language switching and executive functions in bilingualism? Introducing a within group analysis approach. Front Psychol 2011;2:183 [FREE Full text] [doi: 10.3389/fpsyg.2011.00183] [Medline: 21869878]
60. Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A, Wager TD. The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis. Cogn Psychol 2000 Aug;41(1):49-100. [doi: 10.1006/cogp.1999.0734] [Medline: 10945922]
61. Gronwall DM. Paced auditory serial-addition task: a measure of recovery from concussion. Percept Mot Skills 1977 Apr;44(2):367-373. [doi: 10.2466/pms.1977.44.2.367] [Medline: 866038]

62. Rao SM, Leo GJ, Haughton VM, St Aubin-Faubert AP, Bernardin L. Correlation of magnetic resonance imaging with neuropsychological testing in multiple sclerosis. Neurology 1989 Feb;39(2 Pt 1):161-166. [Medline: 2915783]
63. Tombaugh TN. A comprehensive review of the Paced Auditory Serial Addition Test (PASAT). Arch Clin Neuropsychol 2006 Jan;21(1):53-76 [FREE Full text] [doi: 10.1016/j.acn.2005.07.006] [Medline: 16290063]
64. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [Medline: 11556941]
65. Schwarzer R, Jerusalem M. Causal and control beliefs. In: Weinman J, Wright S, Johnston M, editors. Measures in Health Psychology: A Users Portfolio. Windsor, UK: Nfer-Nelson; 1995:35-37.
66. Fann JR, Bombardier CH, Dikmen S, Esselman P, Warms CA, Pelzer E, et al. Validity of the Patient Health Questionnaire-9 in assessing depression following traumatic brain injury. J Head Trauma Rehabil 2005;20(6):501-511. [Medline: 16304487]
67. Cook KF, Bombardier CH, Bamer AM, Choi SW, Kroenke K, Fann JR. Do somatic and cognitive symptoms of traumatic brain injury confound depression screening? Arch Phys Med Rehabil 2011 May;92(5):818-823 [FREE Full text] [doi: 10.1016/j.apmr.2010.12.008] [Medline: 21530731]
68. Hawley L, Gerber D, Morey C. Improving personal self-advocacy skills for individuals with brain injury: a randomized pilot feasibility study. Brain Inj 2017;31(3):290-296. [doi: 10.1080/02699052.2016.1250952] [Medline: 28095051]
69. Brands I, Custers M, van Heugten C. Self-efficacy and quality of life after low-intensity neuropsychological rehabilitation: a pre-post intervention study. NeuroRehabilitation 2017;40(4):587-594. [doi: 10.3233/NRE-171446] [Medline: 28211828]
70. Roth R, Isquith P, Gioia G. Behavior Rating Inventory of Executive Function - Adult Version (BRIEF-A). Lutz, Florida: Psychological Assessment Resources; 2005.
71. Waid-Ebbs JK, Wen PS, Heaton SC, Donovan NJ, Velozo C. The item level psychometrics of the behaviour rating inventory of executive function-adult (BRIEF-A) in a TBI sample. Brain Inj 2012;26(13-14):1646-1657. [doi: 10.3109/02699052.2012.700087] [Medline: 22876936]
72. The World Medical Association (WMA). WMA. 2013. WMA declaration of Helsinki – ethical principles for medical research involving human subjects URL: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/ [WebCite Cache ID 6oAjg5rKp]
73. Maeda Y, Kurokawa T, Sakamoto K, Kitamoto I, Ueda K, Tashima S. Electroclinical study of video-game epilepsy. Dev Med Child Neurol 1990 Jun;32(6):493-500. [Medline: 2365142]
74. Lowe C, Rabbitt P. Test/re-test reliability of the CANTAB and ISPOCD neuropsychological batteries: theoretical and practical issues. Cambridge Neuropsychological Test Automated Battery. International Study of Post-Operative Cognitive Dysfunction. Neuropsychologia 1998 Sep;36(9):915-923. [Medline: 9740364]
75. O'Connor C, Colantonio A, Polatajko H. Long term symptoms and limitations of activity of people with traumatic brain injury: a ten-year follow-up. Psychol Rep 2005 Aug;97(1):169-179. [doi: 10.2466/pr0.97.1.169-179] [Medline: 16279322]
76. Rapoport MJ, McCullagh S, Shammi P, Feinstein A. Cognitive impairment associated with major depression following mild and moderate traumatic brain injury. J Neuropsychiatry Clin Neurosci 2005;17(1):61-65. [doi: 10.1176/jnp.17.1.61] [Medline: 15746484]
77. Kreutzer JS, Seel RT, Gourley E. The prevalence and symptom rates of depression after traumatic brain injury: a comprehensive examination. Brain Inj 2001 Jul;15(7):563-576. [doi: 10.1080/02699050010009108] [Medline: 11429086]
78. Burdea GC, Jain A, Rabin B, Pellosie R, Golomb M. Long-term hand tele-rehabilitation on the PlayStation 3: benefits and challenges. 2011 Presented at: Annual International Conference of the IEEE Engineeing in Biology and Medicine Society, EMBC, 2011; August 30-September 3, 2011; Boston, MA, USA p. 1835-1838. [doi: 10.1109/IEMBS.2011.6090522]
79. Björkdahl A, Akerlund E, Svensson S, Esbjörnsson E. A randomized study of computerized working memory training and effects on functioning in everyday life for patients with brain injury. Brain Inj 2013;27(13-14):1658-1665. [doi: 10.3109/02699052.2013.830196] [Medline: 24131298]
80. Fernández E, Bringas ML, Salazar S, Rodríguez D, García ME, Torres M. Clinical impact of RehaCom software for cognitive rehabilitation of patients with acquired brain injury. MEDICC Rev 2012 Oct;14(4):32-35 [FREE Full text] [Medline: 23154316]
81. Ballesteros S, Toril P, Mayas J, Reales JM, Waterworth JA. An ICT-mediated social network in support of successful ageing. Gerontechnology 2014;13(1):37-46. [doi: 10.4017/gt.2014.13.1.007.00]
82. Vahlo J, Kaakinen JK, Holm SK, Koponen A. Digital game dynamics preferences and player types. J Comput Mediat Comm 2017 Feb 08;22(2):88-103. [doi: 10.1111/jcc4.12181]
83. Rutherford BR, Wall MM, Brown PJ, Choo TH, Wager TD, Peterson BS, et al. Patient expectancy as a mediator of placebo effects in antidepressant clinical trials. Am J Psychiatry 2017 Feb 01;174(2):135-142. [doi: 10.1176/appi.ajp.2016.16020225] [Medline: 27609242]

Abbreviations

BRIEF-A: Behavior Rating Inventory of Executive Function-Adult version
GSE: General Self-efficacy Scale

PASAT: Paced Auditory Serial Addition Test
PHQ-9: Patient Health Questionnaire
QoL: quality of life
RA: research assistant
RCT: randomized controlled trial
TBI: traumatic brain injury
TMT: Trail Making Test
WAIS-IV: Wechsler Adult Intelligence Scale, 4th Edition

Edited by G Eysenbach; submitted 01.03.17; peer-reviewed by J Anguera, A DeSmet, R Kretschmann; comments to author 20.06.17; revised version received 15.11.17; accepted 02.01.18; published 19.03.18.

Please cite as:
Välimäki M, Mishina K, Kaakinen JK, Holm SK, Vahlo J, Kirjonen M, Pekurinen V, Tenovuo O, Korkeila J, Hämäläinen H, Sarajuuri J, Rantanen P, Orenius T, Koponen A
Digital Gaming for Improving the Functioning of People With Traumatic Brain Injury: Randomized Clinical Feasibility Study
J Med Internet Res 2018;20(3):e77
URL: http://www.jmir.org/2018/3/e77/
doi:10.2196/jmir.7618
PMID:29555622

Original Paper

Total Knee Replacement and the Effect of Technology on Cocreation for Improved Outcomes and Delivery: Qualitative Multi-Stakeholder Study

Yasmin van Kasteren[1], PhD; Jill Freyne[2], PhD; M Sazzad Hussain[2], PhD

[1]Adaptive Social and Economic Systems, Land and Water, Commonwealth Scientific and Industrial Research Organization, Dutton Park, QLD, Australia
[2]Australian e-Health Research Centre, Health and Biosecurity, Commonwealth Scientific and Industrial Research Organization, Epping, NSW, Australia

Corresponding Author:
M Sazzad Hussain, PhD
Australian e-Health Research Centre
Health and Biosecurity
Commonwealth Scientific and Industrial Research Organization
GPO Box 76
Epping, NSW, 1710
Australia
Phone: 61 2 9372 4177
Fax: 61 2 9372 4106
Email: sazzad.hussain@csiro.au

Abstract

Background: The growth in patient-centered care delivery combined with the rising costs of health care have perhaps not unsurprisingly been matched by a proliferation of patient-centered technology. This paper takes a multistakeholder approach to explore how digital technology can support the cocreation of value between patients and their care teams in the delivery of total knee replacement (TKR) surgery, an increasingly common procedure to return mobility and relieve pain for people suffering from osteoarthritis.

Objective: The aim of this study was to investigate communications and interactions between patients and care teams in the delivery of TKR to identify opportunities for digital technology to add value to TKR health care service by enhancing the cocreation of value.

Methods: A multistakeholder qualitative study of user needs was conducted with Australian stakeholders (N=34): surgeons (n=12), physiotherapists (n=3), patients (n=11), and general practitioners (n=8). Data from focus groups and interviews were recorded, transcribed, and analyzed using thematic analysis.

Results: Encounters between patients and their care teams are information-rich but time-poor. Results showed seven different stages of the TKR journey that starts with referral to a surgeon and ends with a postoperative review at 12 months. Each stage of the journey has different information and communication challenges that can be enhanced by digital technology. Opportunities for digital technology include improved waiting list management, supporting and reinforcing patient retention and recall of information, motivating and supporting rehabilitation, improving patient preparation for hospital stay, and reducing risks and anxiety associated with postoperative wound care.

Conclusions: Digital technology can add value to patients' care team communications by enhancing information flow, assisting patient recall and retention of information, improving accessibility and portability of information, tailoring information to individual needs, and by providing patients with tools to engage in their own health care management. For care teams, digital technology can add value through early detection of postoperative complications, proactive surveillance of health data for postoperative patients and patients on waiting lists, higher compliance with rehabilitation programs, and reduced length of stay. Digital technology has the potential to improve patient satisfaction and outcomes, as well as potentially reduce hospital length of stay and the burden of disease associated with postoperative morbidity.

(J Med Internet Res 2018;20(3):e95) doi:10.2196/jmir.7541

KEYWORDS

arthroplasty; replacement; osteoarthritis; patient participation; consumer health informatics; technology; telemedicine; rehabilitation; self-care; exercise therapy; human computer interaction; wearables

Introduction

Health care delivery is changing from paternalistic models toward more participative and patient-centered models in which patients, as consumers of health care services, are viewed as active rather than passive participants in their own health care [1,2]. The growth in patient-centered care delivery combined with the rising costs of health care have perhaps not unsurprisingly been matched by a proliferation of patient-centered technology [3-8]. In this research, we explore how care teams can utilize digital technology to support patient-centered care for better outcomes and higher patient satisfaction in total knee replacement (TKR).

TKR is an increasingly popular elective surgery for people wanting relief from knee pain and loss of lifestyle because of reduced functionality. Osteoarthritis is the primary diagnosis of knee replacement patients (accounting for 97% of surgeries) [9]. The average age of patients undergoing TKR in Australia is 69 years [10]. When recommended, TKR is generally effective in reducing knee pain and improving functionality [11,12]; however, despite many successes, there is still a relatively high degree of patient dissatisfaction following TKR. A 2007 survey of 10,000 TKR patients in the United Kingdom showed that only 82% of TKR patients were satisfied with their outcome 12 months after surgery [13]. More recently, a 2012 systematic review of journal articles on postoperative outcomes for TKR indicated that as many as 10% to 34% of TKR patients continue to experience long-term pain after knee replacement [14].

In this paper, we explore the facilitation of patient-centered care using digital technology in a clinical setting for the delivery of TKR. Specifically, we draw on the concept of cocreation of value based on service dominant logic [15-17]. Service dominant logic is a marketing perspective that calls attention to the importance of services in creating value for consumers whereby value, or the utility of a service, is determined by the customer [16]. Although TKR, akin to other health services, does involve the exchange of goods (implant, medication, and consumables), it is predominantly a service provided by hospitals, clinics, and care teams [2,18-20]. Similar to other services, health care requires the active participation of consumers (ie, patients) [18]. The value of the service for the consumer is created, or rather cocreated, with the service care team. Value or utility of a service to customers is uniquely determined by the customer [15]. TKR surgery is a means to an end. The value or utility of the surgery is determined by each individual, and the value for the patient could be about playing or being able to play with grandchildren, going on an overseas trip, or being able to continue to play golf. Interactions between service care teams and consumers are opportunities to add value. The way in which interactions are managed can facilitate (or destroy) consumer value [21-24]. Digital technology has the potential to enhance the delivery of health care services by adding value to the interactions and communications between patients and their care teams [25-28].

In this paper, we explore the TKR journey from start (first referral to a surgeon) to finish (12 months after surgery) to understand the disparate needs of stakeholders and identify opportunities for digital technology to add value to TKR health care service delivery by enhancing the cocreation of value.

Methods

Design: Qualitative Study

Ethics approval was obtained from the *Commonwealth Scientific and Industrial Research Organisation* (CSIRO) Health and Medical Human Research Ethics Committee: low risk review panel. Qualitative data were collected using focus groups and interviews conducted in Queensland and New South Wales, Australia from June 2015 to September 2015. In total, 34 people (patients, general practitioners [GPs], and clinicians) participated in the research. Participants were recruited based on a maximum variation sampling approach [29] to ensure a diversity of opinion and experience. All focus groups, interviews and discussions were audio-recorded and transcribed. Qualitative analysis was carried out using NVivo software (QSR International) to assist in the management of the coding task. Thematic analysis was used to identify emerging themes [30].

Study Population

Participants (N=34) included three different target groups: patients (n=11), GPs (n=8), and clinicians (n=15), defined in this research as surgeons (n=10), research fellows (n=2), and physiotherapists (n=3).

Patients and GPs were invited to attend structured focus groups and clinicians were invited for interviews or discussions. In the following sections, we detail the participant recruitment process and how data were collected. For each group of participants, we outline the key questions. A more comprehensive list of focus groups and interview questions is available in Multimedia Appendix 1.

Methods of Data Collection

Focus Groups

Focus groups (n=17) were used to collect data from GPs and patients. For convenience and expediency, we contracted two different Sydney-based medical market research companies to recruit participants for general practitioner (GP; n=8) and patient focus groups (n=9) because market research companies have access to a large number of GPs and patients with the relevant history of TKR needed to convene a focus group. The focus groups were conducted by CSIRO researchers and lasted around 1 hour long. All participants signed consent forms.

Table 1. General practitioner details.

Demographics	Bulk billing (Medicare), n	Private, n
Gender		
Male	1	3
Female	1	3
Age (years)		
35-44	1	2
45-54	1	2
55+		2
Years' experience		
6-10		1
11-19	2	2
20+		3
Hours worked		
21-40	1	1
41-60		6

GP focus groups (n=8): GPs were screened based on gender, age, years in general practice, weekly patient care hours, proportion of public patients (100% funded by Medicare), and frequency of referral of patients for TKR (see Table 1). The aim of the GP focus group was to understand more about the role of GPs in TKR in Australia because patients cannot consult a surgeon without a referral from a GP. Questions included recounting some of their experiences in referring patients for TKR, with examples of different types of patients, patient outcomes, and differences in referral and treatment between private and public patients. GPs were also asked about any treatments they recommended presurgery, such as exercise or weight loss, as well as their role in postsurgery.

Patient focus groups (n=9): patients were screened based on gender, age, type of knee surgery, principal diagnosis, year of surgery, patient type: private or public, and choice of rehabilitation. Table 2 summarizes patient demographics. A total of 10 patients were recruited, but one did not attend. The aim of the patient focus group was to understand the experience of TKR patients as a basis for developing an app or technology solution to assist them through TKR. Patients were asked about how they came to consider and make the decision to undergo TKR surgery, specifically asking about the process they went through and what sort of information they needed along the way to make decisions or to fulfil their role as a patient. Patients were then asked to describe their experience of TKR surgery, including the hospital stay, the rehabilitation, and recovery from surgery. Patients were also asked about what might have been done differently or better. Because the experience of TKR was uniformly positive among the 9 patients in the focus group, researchers used convenience sampling (word of mouth) to recruit and interview a further 2 patient participants to capture a more balanced view of the patient experience.

Interviews or Discussions

A total of 17 people, including patients (n=2) and clinicians (n=15), participated in semistructured interviews or discussions. Discussions refer to interviews involving more than one participant. Participants were recruited using convenience sampling, people known to either the research team or the research sponsor. The research included both public and private sector practitioners. Participation in interviews or discussions with surgeons and physiotherapists was opportunistic and self-organizing. Participant numbers varied between 1 and 8 attending depending on availability and opportunity on the day. One interview was conducted by phone, all others were face-to-face. The duration of interviews varied from 25 to 120 min dependent on the size of the group.

Patients (n=2): for maximum sample variation, it was important to explore bad patient experience or service failure. Researchers used convenience sampling (word of mouth) to recruit 2 patient participants who were interviewed together. Both had postoperative complications, multiple revisions and knee replacements. The interview covered some of the same ground as the focus group questions but focused on the experience of, and reasons for, service failure.

Clinicians (n=15): clinician interviews or discussions included orthopedic surgeons (OS, n=10), research fellows (n=2), and physiotherapists (n=3). Participants were recruited from both Queensland and New South Wales, see Table 3. Although interview's and discussions followed a broad direction, they were less structured than the focus group because of the multidisciplinary nature of groups, the number of attendees, differences in patient base (private and public, local or remote), as well as interstate differences in hospital practice. The aim of the interviews with clinicians was to understand the process of TKR. Clinicians were asked about each step in the TKR timeline (for both private and public patients) from referral to 12 months post surgery, which is when TKR patients are deemed to have

fully recovered from surgery. Particular emphasis was placed on understanding the communication and information exchange between clinicians and patients. Clinicians were asked to explain how and when information was provided to patients and discuss issues where patients did not fully understand or follow instructions. Clinicians were also asked to describe types of patients, as well as their experience of patient compliance. Finally, surgeons were asked to discuss how they managed patient satisfaction and to describe circumstances leading to poor outcomes.

Table 2. Patient demographics, including both focus group (n=9) and interview participants (n=2).

Demographics	Patient types		
	Private patients, n	Medicare patients, n	Covered by workers compensation, n
Location			
Queensland	1		1
New South Wales	7	1	1
Gender			
Male	4		1
Female	4	1	1
Rehabilitation			
Inpatient rehabilitation hospitals	5		1
Outpatient rehabilitation	1	1	
In-home physiotherapy			1
Local physiotherapy clinic	1		
No formal physiotherapy	1		
Age (years)			
50-59	4		1
60-69	3	1	1
70-79	1		

Table 3. Clinicians' breakdown.

Demographics	Clinicians		
	Surgeons, n	Fellows, n	Physiotherapist, n
Location			
Queensland	7	1	2
New South Wales	3	1	1
Gender			
Male	8	1	2
Female	2	1	1
Health care regime			
Public	2		1
Private		1	2
Public and private	8	1	

Results

Two themes emerged from patient and clinician data on the experience of TKR. First, for patients TKR is not an event but a long journey (12+ months). The journey is clearly marked by stages with different interactions and opportunities for cocreation at each stage. Second, effective communication and information flow between patients and clinicians is a key cocreation task in TKR, irrespective of the stage of the journey.

Theme 1—The Patient Total Knee Replacement Journey

It was clear from the way patients and clinicians described TKR that it is not an event but a journey of 12 months or more, which

officially starts when a patient obtains a referral to an orthopedic surgeon (OS) from their local GP. Seven stages of cocreation were identified: referral, consultation, prehabilitation, perioperative, hospital stay, rehabilitation, and postoperative (Table 4).

1. Referral

In Australia, patients require a referral from a GP to access an OS. The GP referral marks the start of the patient's TKR journey. Most patients will have seen their GPs over the years for treatment and management of the increasing knee pain and loss of functionality typical associated with osteoarthritis before getting a referral for surgery. Some patients anxious to avoid loss of lifestyle request TKR early, others, because of fear of surgery or concerns about the outcomes, leave it too long:

> *Some want it way too early. They just think, oh, a little bit of a twinge, I might get a new knee.* [GP 2]

> *I had another guy the opposite way who put his knee replacement off for ages and ages.* [GP 8]

The referral stage is marked by information seeking and decision making. Patients must decide if they wish to pursue surgery, and private patients can also decide which surgeon they want to be referred to. To support decision making, patients are very active in information seeking, principally using the Internet, GPs recommendations, and increasingly via their social networks. Patients research the medical procedures involved, the experience of TKR, and prospective surgeons. GPs reported that private patients are increasingly assertive in exercising their rights as consumers of health care services. They are actively requesting TKR surgery to avoid loss of lifestyle and/or referral to a specific surgeon or a number of surgeons based on their research:

> *Lots of people I know they've had knee replacements; they were all very successful, so I just took their word, and their recommendations.* [Patient 8]

Table 4. The stages of the patient journey.

Stage	Cocreation network	Patients role	Timelines and details[a]
(1) Referral	GP[b], social network[c]	Information seeking. Decision making: decision to consider surgery.	Private: 2-12 weeks; Public: 12 months.
(2) Consultation	Administrative staff, OS[d]	Understanding information, deciding on surgeon (private patients), risk management, establishment of a relationship of trust with surgeon.	Private: 2-4 weeks; Public: approximately 12 months.
(3) Prehabilitation	Physiotherapist	Being fit for surgery and exercise completion.	Private: more likely to go straight to surgery; Public: emphasis on home exercise therapy.
(4) Perioperative	Administrative staff, occupational therapist, physiotherapist, registered nurse (RN)	Preparation for hospital admission and discharge.	Public and private: 2 to 4 weeks before surgery.
(5) Hospital stay	Administrative staff, OS, RN, physiotherapist, occupational therapist, social network	Following instructions and providing feedback on progress.	Public and private: walking day 1, discharge day 3 to 5.
(6) Rehabilitation	Physiotherapist	Exercise completion and providing feedback on progress.	Private: in- and outpatient rehabilitation and private physiotherapy; Public: outpatient rehabilitation, limited public funding for private physiotherapy.
(7) Postoperative	Administrative staff, occupational therapist, RN, GP	Detecting and reporting complications. Following medication. Staying positive.	Private: scheduled appointments, removal of surgical clips at clinic or by OS at 2 weeks, OS at 6 and 12 weeks and at 12 months; Public: fixed schedule, removal of surgical clips at clinic or GP at 2 weeks, OS at 6 and 12 weeks and at 12 months.
(8) Complications (restart from perioperative)	GP and/or OS		Private: more access to OS over and above scheduled appointments; Public: more gatekeepers.

[a]There is some variation in practice especially in private surgeons, the time lines suggested reflect common practice.

[b]GP: general practitioner.

[c]family and friends.

[d]OS: orthopedic surgeon.

Public patients have less, if any, choice in their surgeon and are therefore more passive consumers at this stage of their TKR journey. The time between referral by a GP and the first consultation with a surgeon varies based on health cover. Private patients will have their first consultation with an orthopedic surgeon usually between 2 to 4 weeks after referral. However, because TKR is an elective surgery, public patients can wait up to a year to see a surgeon. Waiting lists are problematic because GPs try and get public patients onto to a waiting list early so that by the time they need surgery, they will already be on a surgeon's list. This, however, creates additional work for surgeons who must review all patients on their lists yearly. However, these annual reviews are often the only thing assuring patients that they have maintained their position on the waiting list.

2. Consultation

The consultation stage for patients is about risk assessment and forging a relationship of trust with their orthopedic surgeon. Importantly, during this stage, patients establish expectations about the outcomes of TKR, which in turn affects their subsequent satisfaction with the surgery:

> *[S]atisfaction is directly linked with expectations. So if there are those conversations preoperatively about this will take you 12 months or 2 years to get over, I think they are much more willing to accept that that's the case, rather than being dissatisfied for the first 12 months.* [Fellow 1]

In these initial consultations, establishing a relationship with the surgeon is very important. Patients must be comfortable with and have faith in the surgeon's ability to deliver the outcomes they expect:

> *I went to a surgeon first, someone who had done my father's knee...he wasn't the surgeon I ended up with because I couldn't stand the guy. I wasn't going to let him cut me open.* [Patient 1]

3. Prehabilitation

Once a decision to undergo surgery is taken, private patients have a 2 to 8 week wait for their surgery to be scheduled. Private surgeons tend not to recommend prehabilitation:

> *Very rarely would someone of mine go in for prehab[ilitation]...I don't know if I'm right or wrong—but I tend to reserve the rehab[ilitation] for postsurgery.* [Surgeon 4]

In the public sector, waiting time can be over 12 months. GPs, physiotherapists, surgeons, and public services use this time to promote prehabilitation:

> *If they come in at a reasonable period before their surgery date, for instance, if they've got, say, 4 to 6 weeks beforehand, it gives us the opportunity to possibly improve, say, their knee extension prior to surgery, because we know that will dramatically improve their walking ability postop.* [Physiotherapist 1]

> *I don't think you will come to a knee replacement for some time and what I'm going to do is plan to see you in a year but I want you to do [the exercises] and I'm going to write to the GP and say if there's any major changes or a problem I'm happy to see you earlier. So they've got a prehabilitation process.* [Surgeon 10]

4. Perioperative

Once surgery is scheduled, the perioperative stage begins. This stage is information rich. Hospitals and care teams need to communicate with and collect information from the patient with a view to facilitating a cost-efficient hospital experience and a timely admission and discharge.

5. Hospital Stay

Patients admitted to hospital have their surgery and are typically discharged 3 and 5 days after surgery. Private patients in large metropolitan areas have the option of being discharged to inpatient rehabilitation clinics for a 1 or 2 week stay where inpatient physiotherapy is delivered. Public patients who have less access to rehabilitation services post surgery may be kept in the hospital longer if the surgeon feels they might benefit from a few days extra of inpatient physiotherapy:

> *The big difference between public and private, as far as how long they stay in hospital, is really dictated by access to rehab. Private patients can get into rehab, whereas public patients...rehab is almost impossible to find...at least they're receiving some inpatient physio...before going home.* [Surgeon 2]

6. Rehabilitation

Rehabilitation begins in the hospital and continues into the postoperative stage. In-patient physiotherapy starts on the day after surgery when patients are expected to start walking with aids. In addition to walking, patients also have to complete exercises designed to improve knee extension (straightening), range of movement (bending), and function (walking, steps, and running). Rehabilitation continues after discharge at inpatient or outpatient clinics, or with local physiotherapy services. Proximity to rehabilitation facilities or physiotherapists and private health care cover determine access to rehabilitation. However, all physiotherapists recommend in-home exercise and self-management of prescribed in-home exercise as an important part of rehabilitation.

7. Postoperative

Patients will have their wound assessed and clips removed 2 weeks after surgery by their GP in a clinic or by their surgeon. They also typically see their surgeon at 6 weeks to check for wound infection and at 12 months to assess recovery and the performance of the prosthetic. In the postoperative stage, the patient is responsible for detecting and reporting postsurgical complications, including wound infection, clots, and deep vein thrombosis. Although the risks of complications post surgery are low, the consequences are very high, which can cause some uncertainty and anxiety in patients:

> *Well, with things like that, a lot of patients at 6 weeks are worried, "My knee is hot and red." And [knees post surgery] are all hot and red.* [Surgeon 1]

Theme 2—Information and Communication Flow

Two-way communication between patients and care teams is essential in the TKR journey. Clinicians, including surgeons, physiotherapists, and others, need to communicate information effectively to manage patient expectations. Communication is also essential to ensure informed consent and decision making and to ensure patients meet the surgeon's and hospital's expectations for self-management pre and post surgery. Equally, patients need to communicate information effectively to care teams for accurate diagnosis and treatment in the consultation and postoperative phases. Our analysis revealed issues with information flow, gaps in communication, and communication problems.

Information Flow

Care teams and hospitals have to communicate the same information to patients repeatedly. This requires efficiency in communication. Due to the limited amount of face-to-face time, printed information is the principle means of supporting information communication to patients at all stages of their TKR journey:

> *When every patient books for an operation, we give them [an information pack] this is a preloaded what paperwork is required, the consent form, the hospital admission paperwork, thing about knee replacement et cetera, et cetera.* [Surgeon 5]

> *I got a booklet…and [the surgeon] explained a little bit also with an instrument of some sort; I don't know. Showed how the joints fit together, and he gave me the brochure and sent me off.* [Patient 4]

Increasingly hospitals are organizing presurgery education or assessment sessions during the perioperative phase. These sessions often combine group information sessions and one-on-one consultations with nurses, physiotherapists, and occupational therapists:

> *Well, I have worked in two private hospitals, and they had significant differences in their length of stay. The big difference has been the patients who have been preoped have everything organised, arranged, and mostly their expectation of discharge. So they come in and they go, I'm having a total knee replacement, I'm going to be in, my surgery is on Monday, I'm going to be home Friday.* [Physiotherapist 3]

However, interview data revealed that patients do not necessarily absorb, retain, or act on the information provided:

> *We always inform the patients not to take antibiotics. But I do the review at 2 weeks; a lot of people are full of antibiotics for unnecessary reasons.* [Surgeon 1]

> *There was one patient that will forever stick in my mind. She was in absolute tears when I called her, and she was distraught...I found out that she...thought that she had to stay in the house. [S]he thought that she was trapped in the house until she went back to see the GP at 2 weeks and the surgeon at 6 weeks.* [Physiotherapist 2]

Information can also allay fears and concerns:

> *We often have to ask them, have you done the [presurgery] education? And you know the ones that haven't, because...they're apprehensive about getting out of bed.* [Physiotherapist 2]

Physiotherapists also provide patients with a lot of information. Physiotherapist's explain to patients how to modify everyday activities to cope with limited functionality (eg, how to climb stairs safely and the best way to get in and out of bed) and prescribe exercise plans which include the type of exercise, the number of repetitions, and the recommended frequency. Importantly, they also teach patients to perform exercises correctly. In addition to providing information, physiotherapists also have to motivate the patient to complete their exercise program:

> *So a lot of them get given information sheets and stuff like that, they watch videos in there as well and then get given a whole lot of stuff to do at home, so I think a good rehab is a rehab that will encourage them to be self-motivated rather than just doing 4 hours of rehab a week while at rehab.* [Physiotherapist 1]

Patients must also convey information to care teams. Although patients are not medical experts, they are the only ones who can communicate what they feel and experience, therefore giving them *expert* knowledge. Surgeons and physiotherapists need quality feedback from patients for their diagnosis and ongoing care. Pain and sleep are two important variables that help surgeons and physiotherapists assess progress. Sleep quality is directly associated with pain, and care teams use both to understand patient progress. Other key information patients are required to provide to the doctor is physical functionality:

> *I think it's critical…to take a history about physical status, what are they doing and I [cover] hills and stairs, crouches and squats, kneeling and ladders, running and jumping,…the way you answer me will tell me where in the knee the problem lies…each of those activities, and the way you weight them helps me to know where the problem is.* [Surgeon 10]

> *The decision to operate is—it's their pain level, it's their function, it's the, can you walk—I can't walk around the shops, I can't sleep…and my usual statement to patients is, "You have your operation when you are ready."* [Surgeon 1]

Cocreation Styles

In addition to the cocreation tasks associated with information gathering and decision making, patient's personalities and life experience mean that they establish different types of relationships with their surgeons and physiotherapists. The type of relationship they establish is particularly apparent in the postoperative stage, where lack of expertise or experience and the high cost of undetected complications makes patients anxious and fearful:

> *Rare complications when you have them, they're disastrous (64.10) infected prosthesis, it's terrible. I've had patients who've had a protracted postoperative of 6 months said, I wish I never set my eyes on that surgeon.* [GP 3]

> *Because if you wash [infected joints] out in the first few weeks, you may be able to save the joint. So it actually matters a lot. And the main reason, probably we see the thousands [of patients at] 6 weeks is to pick up the one [infection].* [Surgeon 1]

Public and private patients generally see their surgeons at 6 weeks and at 12 weeks. In the postoperative stage, private patients have more access to their surgeons. They are more readily able, and perhaps feel more entitled, to communicate directly with their surgeon:

> *There's usually a heightened sense of awareness and patients are paranoid about clots and infection, so most people will ring at the drop of a hat if they're worried...if someone's worried I'd rather see them.* [Surgeon 9]

Public patients are generally encouraged to see their GP if they have concerns about postsurgical complications, despite a number of surgeons reporting that GPs lacked knowledge to correctly treat postsurgical infections. Over and above differences in health care cover, differences in personality and communications style result in both under- and overreporting.

Overreporting occurs when patients are overanxious and frequently contact the surgeons directly or via the practice receptionist for private patients or via the GP for public patients. Clinician time is limited, so to relieve pressure on care teams and GPs, practice receptionists can act as gatekeepers, filtering requests. On rare occasions, their lack of clinical expertise, coupled with patient's limited ability to communicate their condition, could result in genuine complications being overlooked:

> *When I go in there the message that [my surgeon] had [from his receptionist] was that I had nicked the top of the scar. It was wide open, I had to have a tourniquet to stop the bleeding...that was the arrogance of a front line staff member who had not even passed the message on.* [Patient 11]

Underreporting also occurs and is equally problematic. Patients who do not want to bother the surgeon will wait until the next appointment to raise concerns that again can exacerbate outcomes:

> *What I say to people is..."See you in 6 weeks." And then they sit there for 2 weeks with pus pouring out of their knee, because they are seeing the doctor at 6 weeks...so whatever happens in between, they just hold on to it.* [Surgeon 1]

Most patients tend to follow medical advice. Many patients referred to themselves as a "good patient," following care team advice and instructions, in particular with respect to medication and exercise:

> *I think I was an excellent patient...so I did all the things I was told to do [in rehab] because I just wanted the best recovery I could have and I just thought their exercises, their knee clinic I think they called it, it was just really good.* [Patient 3]

Reasons for not completing rehabilitation or home exercise included access to rehabilitation, having to rely on others to drive them to a clinic, the need to return to work, and pain:

> *The rehab, getting that bend in the knee, I can still remember tears streaming down my face trying to do what they were telling me to do, to get that particular degree that they wanted.* [Patient 3]

> *I was really concentrating on exercising and that, while I was in the hospital…then I wanted to get back to work, because I was totally bored. So I didn't do as much rehabilitation as what I should be doing.* [Patient 7]

Due to pain and difficulty of completing exercises, some patients (Patient 11) developed very negative attitudes toward their physiotherapists describing them as "overzealous, rude, and condescending."

Discussion

Principal Findings

This research explored TKR patient-clinician interactions looking for opportunities for digital technology to enhance cocreation and add value at different stages of the TKR journey. Technology can add value by enhancing and increasing opportunities for two-way communication between clinicians and their patients.

Communication From Clinician to Patient

Results showed that interactions between clinicians and patients are time-poor but information-rich, and patients do not necessarily retain or recall the information or instructions given to them by their care team because information given by clinicians, and in particular surgeons, to their patients is often technical and complex [31]. About half a surgeon's consultation time is spent explaining the medical condition, treatment options, and surgical procedure; a good deal of which is required to meet legal obligations [31]. Although this information is necessary for patients to understand their condition, the risks associated with treatment options, and to provide informed consent to treatment plans [31-34], it has been demonstrated that patients have very poor recall of information provided to them by clinicians, especially if, as with TKR, they are older, in pain, or anxious [35]. Patients also have selective information retention and generally have higher recall of information about the diagnosis than about the treatment options [35]. Similarly, with rehabilitation, patient recall of exercises, sets, and repetitions is particularly problematic for older adults, and recall is not substantially improved by the provision of a memory sheet [36]. Results showed a clear role for digital technology to add value through improved communication and information flow between clinicians and patients. Digital technology can facilitate the following:

- *Deliver the right information at the right time*. Mobile phone features such as short message service, push notifications, reminders, tasks, and alarms are well suited to supporting real-time information delivery and "just-in-time" access [4,8]. The results clearly identified different information needs at different stages of TKR.

Technology-supported information delivery can deliver the right information at the right time, making it both easier to absorb and readily accessible to add value through improved patient recall and compliance. Reminders have been successful in increasing adherence for routine daily tasks such as medication management [4]. Mobile notifications may also increase adherence to home-based physiotherapy, as research has shown that setting regular times for exercise and integrating exercise into routine is more likely to encourage successful completion [37]. Interactive tasks or checklists could potentially help patients prepare for hospital admission, complete rehabilitation plans, and take medication as prescribed.

- *Convey complex information in a more engaging way, including text, imagery, audio, or video* [36], thereby adding value through better patient understanding of treatment and rehabilitation. Mobile phone apps with videos could support demonstrating the surgery and the treatment options, and video is a great medium for providing exercise demonstrations to support safe and correct exercise at home [38-42].
- *Enhance motivation and compliance with exercise programs through interactive gaming and rewards systems.* Gamification is a commonly used engagement strategy to increase motivation by providing positive feedback to users, which in TKR could include compliance with physiotherapy and achievement of functionality milestones. Achieving functionality milestones could support patients' sense of positive progress toward recovery.
- *Increase the number of interactions (albeit indirect) and thereby the opportunities for adding value between care team and patients*. Encounters and interactions with clinicians are dictated by standardized care pathways that determine duration and number of appointments. Push notifications and just-in-time delivery of information can add value by indirectly increasing the communication or interactions from care team to patients between appointments. Mobile phone apps can also provide patients who cannot or do not have access to rehabilitation services with support in completing their rehabilitation. A mobile phone app for cardiac rehabilitation has been proven to increase uptake, adherence, and completion of patient rehabilitation for patients who have had a heart attack in comparison to face-to-face outpatient clinic [39,40].
- *Tailor information to individual needs to deliver personalized solutions.* Mobile technology can ensure that individual patients receive personally relevant information and/or can make choices in how they receive the information and which support tools they use. Mobile phone apps can also potentially match different behavioral interventions to different cocreation styles to optimize outcomes [24].

Communication From Patient to Clinician

Patients have an *expert* role in that only they can describe what they are feeling and experiencing. This patient input is a crucial element of clinical diagnosis and treatment [32,43,44]. Results show that technology could add value to patient clinician communication by supporting patients in their role of an *expert* to better and more accurately communicate how they are feeling and what they are experiencing. Self-reported information can be difficult for patients to recall, inconsistent over time, and difficult to convey [45]. Research shows that patient self-reports on sleep and pain are unreliable. For participants with fair or poor health, no correlation was observed between subjective and objective sleep measures [46], and patients with a high social desirability bias (desire to be viewed favorably by their surgeon) report higher level of pain [47]. For patient clinician communication, digital technology can facilitate the following:

Enhance the quality and frequency of patient communication with their care team using self-monitoring tools and wearables. Self-monitoring using mobile phones could assist patients in recalling the timing of events and improve accuracy and validity of self-reports on pain or recovery progress with text-based diaries, photo or video blogs, tools to track pain, medication use, mood, physiotherapy, knee range of motion, and more. This type of reporting could be further enhanced with the use of wearable technologies such as activity trackers, body worn sensors, or Bluetooth-enabled thermometers or scales. By linking data from digital technology to a Web portal accessible by the care team, daily progress data could be used to flag patients not progressing as expected, allowing clinicians to move appointments forward as needed. Remote monitoring of progress could potentially increase motivation and compliance as patients extend the role of *good patient* to their self-management tasks, knowing that they are been observed. This approach could also be used for waiting list management to supplement the yearly orthopedic consult.

Assist in the early detection of postoperative complications and reduce patient anxiety. Results showed that the most critical time for effective patient communication is the postoperative period (0-12 weeks post surgery) because even though the risk of infection and complication after TKR are low, the resulting morbidity is high, and patients are naturally anxious. With only limited access to surgeons (3 postoperative visits in 12 weeks), it is largely the patient's responsibility to detect and report postoperative complications and infections. Results showed that patients both under- and over-report postsurgical complications. There is very clearly a role for technology to add value in the detection of postsurgical complications to help reduce patient anxiety and uncertainty and improve patient outcomes. Solutions might include telehealth services for the 12-week postoperative period or intelligent technology such as smart wound dressings, wearable devices, and heat detectors for infections [48,49].

Strengths and Limitations

The strength of this research is that it provides a multistakeholder perspective on how care teams can use digital technology to add value to patient clinician interactions and in so doing, potentially improve patient experience and satisfaction. Limitations of this research include the limited generalizability of qualitative data. We also acknowledge an overrepresentation of the private sector perspective (patients and care teams), even though 70% of TKR in Australia occurs in private hospitals [50]. Although research was conducted in two major cities, Sydney and Brisbane, this study is not necessarily representative of TKR journey in whole of Australia, as only 60% of TKR

surgeries occur in capital cities [51]. Focus groups conducted in remote and less affluent areas may have yielded different patient experiences because access to TKR and physiotherapy are affected by proximity to hospitals and rehabilitation facilities, as well as affordability. Finally, this research was cofunded by a commercial partner who has engaged CSIRO to conduct this research independently on their behalf. CSIRO was solely responsible for decisions on the study design, analysis, and interpretation of data. This research has led to the design and development of a digital orthopedic rehabilitation platform, which is being evaluated through a multihospital randomized controlled trial and registered with the Australian New Zealand Clinical Trials Registry (ACTRN12616000504415) [52].

Conclusions

Digital technology has the potential to enhance the current model of care for TKR, adding value for all stakeholders through increased and improved communication and information flow. For patients, digital technology could enhance information retention and recall, support the patient as an *expert*, reduce anxiety in the postoperative stage, improve recovery through improved adherence to rehabilitation, and increase satisfaction through supporting personal agency and perceived control. For clinicians, digital technology can enhance the communication and information flow between care teams and patients to improve patient compliance, outcomes, and satisfaction. For health care providers, digital technology can assist in managing waiting lists, reduce length of stay because patients are better prepared for admission and discharge, and reduce morbidity and burden of disease through early detection of postoperative complications. Digital technology could also potentially reduce the cost of service delivery [53] without compromising patients' outcomes. Although *patient-centered* care is by definition not *technology-centered* care [54], technology can nevertheless assist in the delivery, reinforcement, and accessibility of information if the focus of digital technology design and development (eg, apps) is value-driven and patient-centric [25].

Acknowledgments

The project is cofunded by Johnson and Johnson Medical Devices Australia and CSIRO. The commercial partner, Johnson and Johnson Medical Devices Australia has engaged CSIRO to conduct this research on their behalf and has approved the manuscript for publication. The authors would like to thank Angela Coble, Ben Travers, Julian Ryall, and Nathan Edgington from Johnson and Johnson Medical Devices Australia for their input and contribution. Present address of YVK: Flinders Digital Health Research Centre, Flinders University, GPO Box 2100 Adelaide SA 5001. Yasmin.vankasteren@flinders.edu.au.

Authors' Contributions

YVK designed the study methodology and conducted the thematic analysis. All the authors participated in interviews and focus groups and contributed to the writing of the report.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Interview questions.

[PDF File (Adobe PDF File), 39KB - jmir_v20i3e95_app1.pdf]

References

1. Dunston R, Lee A, Boud D, Brodie P, Chiarella M. Co-production and health system reform: from re-imagining to re-making. Aust J Public Mark 2009;68(1):39-52. [doi: 10.1111/j.1467-8500.2008.00608.x]
2. Janamian T, Crossland L, Wells L. On the road to value co-creation in health care: the role of consumers in defining the destination, planning the journey and sharing the drive. Med J Aust 2016 Apr 18;204(7 Suppl):S12-S14. [Medline: 27078786]
3. Boulos MN, Wheeler S, Tavares C, Jones R. How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. Biomed Eng Online 2011 Apr 05;10(1):24 [FREE Full text] [doi: 10.1186/1475-925X-10-24] [Medline: 21466669]
4. 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]
5. Lewis TL, Wyatt JC. mHealth and mobile medical Apps: a framework to assess risk and promote safer use. J Med Internet Res 2014 Sep 15;16(9):e210 [FREE Full text] [doi: 10.2196/jmir.3133] [Medline: 25223398]
6. Quinn CC, Clough SS, Minor JM, Lender D, Okafor MC, Gruber-Baldini A. WellDoc mobile diabetes management randomized controlled trial: change in clinical and behavioral outcomes and patient and physician satisfaction. Diabetes Technol Ther 2008 Jun;10(3):160-168. [doi: 10.1089/dia.2008.0283] [Medline: 18473689]
7. Terry M. Medical apps for smartphones. Telemed J E Health 2010;16(1):17-22. [Medline: 20070172]

8. Donker T, Petrie K, Proudfoot J, Clarke J, Birch MR, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
9. Peat G, McCarney R, Croft P. Knee pain and osteoarthritis in older adults: a review of community burden and current use of primary health care. Ann Rheum Dis 2001 Feb;60(2):91-97 [FREE Full text] [Medline: 11156538]
10. Graves S. Hip and Knee Arthroplasty: Annual Report. Adelaide: Australian Orthopaedic Association; 2015.
11. Russell TG, Buttrum P, Wootton R, Jull GA. Internet-based outpatient telerehabilitation for patients following total knee arthroplasty: a randomized controlled trial. J Bone Joint Surg Am 2011 Jan 19;93(2):113-120. [doi: 10.2106/JBJS.I.01375] [Medline: 21248209]
12. Pozzi F, Snyder-Mackler L, Zeni J. Physical exercise after knee arthroplasty: a systematic review of controlled trials. Eur J Phys Rehabil Med 2013 Dec;49(6):877-892 [FREE Full text] [Medline: 24172642]
13. Baker PN, van der Meulen JH, Lewsey J, Gregg P, National Joint Registry for EnglandWales. The role of pain and function in determining patient satisfaction after total knee replacement. Data from the National Joint Registry for England and Wales. J Bone Joint Surg Br 2007 Jul;89(7):893-900 [FREE Full text] [doi: 10.1302/0301-620X.89B7.19091] [Medline: 17673581]
14. Beswick AD, Wylde V, Gooberman-Hill R, Blom A, Dieppe P. What proportion of patients report long-term pain after total hip or knee replacement for osteoarthritis? A systematic review of prospective studies in unselected patients. Br Med J Open 2012;2(1):e000435 [FREE Full text] [doi: 10.1136/bmjopen-2011-000435] [Medline: 22357571]
15. Vargo SL, Lusch RF. Service-dominant logic: continuing the evolution. J Acad Mark Sci 2007 Aug 1;36(1):1-10. [doi: 10.1007/s11747-007-0069-6]
16. Vargo SL, Lusch RF. Evolving to a new dominant logic for marketing. J Mark 2004 Jan;68(1):1-17.
17. Lusch RF, Vargo SL. Service-dominant logic: reactions, reflections and refinements. Mark Theory 2006;6(3):281-288. [doi: 10.1177/1470593106066781]
18. Prahalad CK, Ramaswamy V. Co-opting customer competence. Harv Bus Rev 2000;78(1):79-90 [FREE Full text]
19. Berry LL, Bendapudi N. Health care a fertile field for service research. J Serv Res 2007 Nov;10(2):111-122. [doi: 10.1177/1094670507306682]
20. Porter ME. What is value in health care? N Engl J Med 2010 Dec 23;363(26):2477-2481. [doi: 10.1056/NEJMp1011024] [Medline: 21142528]
21. Grönroos C. Value co-creation in service logic: a critical analysis. Mark Theory 2011 Oct;11(3):279-301 [FREE Full text]
22. Prahalad CK, Ramaswamy V. Co‐creating unique value with customers. Strategy Leadership 2004 Jun;32(3):4-9. [doi: 10.1108/10878570410699249]
23. Osei-Frimpong K, Wilson A, Owusu-Frimpong N. Service experiences and dyadic value co-creation in healthcare service delivery: a CIT approach. J Serv Theory and Pract 2015 Jul 13;25(4):443-462. [doi: 10.1108/JSTP-03-2014-0062]
24. McColl-Kennedy JR, Vargo SL, Dagger TS, Sweeney JC, van Kasteren Y. Health care customer value cocreation practice styles. J Serv Res 2012 May 25;15(4):370-389. [doi: 10.1177/1094670512442806]
25. van Limburg M, van Gemert-Pijnen JE, Nijland N, Ossebaard HC, Hendrix RM, Seydel ER. Why business modeling is crucial in the development of eHealth technologies. J Med Internet Res 2011 Dec 28;13(4):e124 [FREE Full text] [doi: 10.2196/jmir.1674] [Medline: 22204896]
26. Loane SS, Webster CM, D'Alessandro S. Identifying consumer value co-created through social support within online health communities. J Macromarketing 2014;35(3):353-367. [doi: 10.1177/0276146714538055]
27. Caridà A, Colurcio M, Melia M. Rethinking and improving the health care service through interactive web technologies. In: Baglieri E, Karmarkar U, editors. Managing Consumer Services: Factory Or Theater?. Cham: Springer; 2014:191-210.
28. Jing Z, Tao W, Xiucheng F. Patient value co-creation in online health communities: social identity effects on customer knowledge contributions and membership continuance intentions in online health communities. J Serv Manage 2015;26(1):72-96 [FREE Full text]
29. Onwuegbuzie AJ, Collins KM. A typology of mixed methods sampling designs in social science research. Qual Rep 2007;12(2):281-316 [FREE Full text]
30. Guest G, MacQueen KM, Namey EE. Applied Thematic Analysis. Thousand Oaks, CA: SAGE Publications; 2011.
31. Levinson W, Chaumeton N. Communication between surgeons and patients in routine office visits. Surgery 1999 Feb;125(2):127-134. [Medline: 10026744]
32. Lazare A, Putnam SM, Lipkin Jr MJ. Three functions of the medical interview. In: The Medical Interview. NY: Springer; 1995:3-19.
33. Rosenberg EE, Lussier MT, Beaudoin C. Lessons for clinicians from physician-patient communication literature. Arch Fam Med 1997;6(3):279-283. [Medline: 9161356]
34. Ha JF, Longnecker N. Doctor-patient communication: a review. Ochsner J 2010;10(1):38-43. [Medline: 21603354]
35. Jansen J, van Weert J, van der Meulen N, van Dulmen S, Heeren T, Bensing J. Recall in older cancer patients: measuring memory for medical information. Gerontologist 2008 Apr;48(2):149-157. [Medline: 18483427]
36. Smith J, Lewis J, Prichard D. Physiotherapy exercise programmes: are instructional exercise sheets effective? Physiother Theory Pract 2005;21(2):93-102. [Medline: 16392462]

37. Campbell R, Evans M, Tucker M, Quilty B, Dieppe P, Donovan J. Why don't patients do their exercises? Understanding non-compliance with physiotherapy in patients with osteoarthritis of the knee. J Epidemiol Community Health 2001 Feb;55(2):132-138 [FREE Full text] [Medline: 11154253]
38. Schooley B, San Nicolas-Rocca T, Burkhard R. Patient-provider communications in outpatient clinic settings: a clinic-based evaluation of mobile device and multimedia mediated communications for patient education. JMIR Mhealth Uhealth 2015 Jan 12;3(1):e2 [FREE Full text] [doi: 10.2196/mhealth.3732] [Medline: 25583145]
39. Varnfield M, Karunanithi M, Lee CK, 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]
40. Varnfield M, Karunanithi MK, Särelä A, Garcia E, Fairfull A, Oldenburg BF, et al. Uptake of a technology-assisted home-care cardiac rehabilitation program. Med J Aust 2011 Feb 21;194(4):S15-S19. [Medline: 21401482]
41. Kingston G, Gray MA, Williams G. A critical review of the evidence on the use of videotapes or DVD to promote patient compliance with home programmes. Disabil Rehabil Assist Technol 2010 May;5(3):153-163. [doi: 10.3109/17483101003671709] [Medline: 20205545]
42. Weeks DL, Brubaker J, Byrt J, Davis M, Hamann L, Reagan J. Videotape instruction versus illustrations for influencing quality of performance, motivation, and confidence to perform simple and complex exercises in healthy subjects. Physiother Theory Pract 2002;18(2):65-73.
43. Tattersall RL. The expert patient: a new approach to chronic disease management for the twenty-first century. Clin Med (Lond) 2002;2(3):227-229. [Medline: 12108472]
44. Chang HJ, Mehta PS, Rosenberg A, Scrimshaw SC. Concerns of patients actively contemplating total knee replacement: differences by race and gender. Arthritis Rheum 2004 Feb 15;51(1):117-123 [FREE Full text] [doi: 10.1002/art.20073] [Medline: 14872464]
45. Hadjistavropoulos T, Craig KD. A theoretical framework for understanding self-report and observational measures of pain: a communications model. Behav Res Ther 2002 May;40(5):551-570. [Medline: 12038648]
46. Lauderdale DS, Knutson KL, Yan LL, Liu K, Rathouz PJ. Self-reported and measured sleep duration: how similar are they? Epidemiology 2008 Nov;19(6):838-845 [FREE Full text] [doi: 10.1097/EDE.0b013e318187a7b0] [Medline: 18854708]
47. Deshields TL, Tait RC, Gfeller JD, Chibnall JT. Relationship between social desirability and self-report in chronic pain patients. Clin J Pain 1995 Sep;11(3):189-193. [Medline: 8535037]
48. Dargaville TR, Farrugia BL, Broadbent JA, Pace S, Upton Z, Voelcker NH. Sensors and imaging for wound healing: a review. Biosens Bioelectron 2013 Mar 15;41:30-42. [doi: 10.1016/j.bios.2012.09.029] [Medline: 23058663]
49. Mehmood N, Hariz A, Templeton S, Voelcker NH. A flexible and low power telemetric sensing and monitoring system for chronic wound diagnostics. Biomed Eng Online 2015 Mar 01;14:17 [FREE Full text] [doi: 10.1186/s12938-015-0011-y] [Medline: 25884377]
50. Australian Orthopaedic Association National Joint Replacement Registry. Adelaide: Australian Orthopaedic Association; 2015. Lay Summary 2015 Annual Report Hip and Knee replacement URL: https://aoanjrr.sahmri.com/documents/10180/217647/Lay%20Summary%202015%20Annual%20Report%20Hip%20and%20Knee%20Replacement [WebCite Cache ID 6xWL0l97u]
51. Australian Institute of Health and Welfare. 2011. Population differences in health-care use for arthritis and osteoporosis in Australia URL: https://www.aihw.gov.au/getmedia/68c55ff6-6ea9-4266-afa2-b4e560c616ef/11835.pdf.aspx?inline=true [WebCite Cache ID 6xWLFVwY8]
52. Hussain MS, Li J, Brindal E, van Kasteren Y, Varnfield M, Reeson A, et al. Supporting the delivery of total knee replacements care for both patients and their clinicians with a mobile app and web-based tool: randomized controlled trial protocol. JMIR Res Protoc 2017 Mar 01;6(3):e32 [FREE Full text] [doi: 10.2196/resprot.6498] [Medline: 28249832]
53. Lusch RF, Vargo SL, Tanniru M. Service, value networks and learning. J Acad Market Sci 2009 Jan 29;38(1):19-31. [doi: 10.1007/s11747-008-0131-z]
54. Stewart M. Towards a global definition of patient centred care. Br Med J 2001 Feb 24;322(7284):444-445 [FREE Full text] [Medline: 11222407]

Abbreviations

GP: general practitioner
OS: orthopedic surgeon
TKR: total knee replacement
CSIRO: Commonwealth Scientific and Industrial Research Organisation

Edited by G Eysenbach; submitted 24.02.17; peer-reviewed by M Sjölinder, N Bashi; comments to author 13.04.17; revised version received 27.08.17; accepted 18.11.17; published 20.03.18.

Please cite as:
van Kasteren Y, Freyne J, Hussain MS
Total Knee Replacement and the Effect of Technology on Cocreation for Improved Outcomes and Delivery: Qualitative Multi-Stakeholder Study
J Med Internet Res 2018;20(3):e95
URL: http://www.jmir.org/2018/3/e95/
doi:10.2196/jmir.7541
PMID:29559424

Review

Ethical Concerns of and Risk Mitigation Strategies for Crowdsourcing Contests and Innovation Challenges: Scoping Review

Joseph D Tucker[1,2,3,4], MD, PhD; Stephen W Pan[1,2,5], PhD; Allison Mathews[2,4], PhD; Gabriella Stein[2,4]; Barry Bayus[6], PhD; Stuart Rennie[7], PhD

[1]University of North Carolina Project-China, Guangzhou, China
[2]Social Entrepreneurship to Spur Health, Guangzhou, China
[3]Faculty of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom
[4]Institute of Global Health and Infectious Diseases, University of North Carolina, Chapel Hill, NC, United States
[5]Department of Public Health, Xi'an Jiaotong-Liverpool University, Suzhou, China
[6]Kenan-Flagler School of Business, University of North Carolina, Chapel Hill, NC, United States
[7]Social Medicine Department, University of North Carolina, Chapel Hill, NC, United States

Corresponding Author:
Joseph D Tucker, MD, PhD
University of North Carolina Project-China
2 Lujing Road
Guangzhou,
China
Phone: 86 13560294997
Email: jdtucker@med.unc.edu

Abstract

Background: Crowdsourcing contests (also called innovation challenges, innovation contests, and inducement prize contests) can be used to solicit multisectoral feedback on health programs and design public health campaigns. They consist of organizing a steering committee, soliciting contributions, engaging the community, judging contributions, recognizing a subset of contributors, and sharing with the community.

Objective: This scoping review describes crowdsourcing contests by stage, examines ethical problems at each stage, and proposes potential ways of mitigating risk.

Methods: Our analysis was anchored in the specific example of a crowdsourcing contest that our team organized to solicit videos promoting condom use in China. The purpose of this contest was to create compelling 1-min videos to promote condom use. We used a scoping review to examine the existing ethical literature on crowdsourcing to help identify and frame ethical concerns at each stage.

Results: Crowdsourcing has a group of individuals solve a problem and then share the solution with the public. Crowdsourcing contests provide an opportunity for community engagement at each stage: organizing, soliciting, promoting, judging, recognizing, and sharing. Crowdsourcing poses several ethical concerns: organizing—potential for excluding community voices; soliciting—potential for overly narrow participation; promoting—potential for divulging confidential information; judging—potential for biased evaluation; recognizing—potential for insufficient recognition of the finalist; and sharing—potential for the solution to not be implemented or widely disseminated.

Conclusions: Crowdsourcing contests can be effective and engaging public health tools but also introduce potential ethical problems. We present methods for the responsible conduct of crowdsourcing contests.

(J Med Internet Res 2018;20(3):e75) doi:10.2196/jmir.8226

KEYWORDS

crowdsourcing; health communication; ethical analysis

Introduction

Crowdsourcing refers to "the practice of obtaining information or services by soliciting input from a large number of people, typically via the internet and often without offering compensation." [1] The term encompasses a wide range of practices that were originally developed to iteratively improve commercial products based on crowd input and to change the traditional relationship between a business and a client [2]. For example, the online encyclopedia Wikipedia allows anonymous volunteers to write, edit, and manage online encyclopedia entries. Wikipedia has rapidly grown and now has 4.9 million articles that are being edited by 70,000 active contributors [3]. Crowdsourcing is used in the government and nonprofit sectors to generate innovative concepts and designs [4]. Crowdsourcing can take a wide variety of forms, including online games [5], distributed health system platforms [6], and contests to solicit new ideas [4].

Our discussion of crowdsourcing will focus on contests, also called innovation challenges, innovation contests, and inducement prize contests. Crowdsourcing contests include prize-based open contests in which individuals or teams work alone and those in which individuals work together. Contests typically include the following stages: organizing a steering committee, soliciting contributions, promoting the contest, judging contributions by experts or the crowd, recognizing excellent contributions, and sharing contributions. In the past 10 years, contests have been used to promote public health [4]. Crowdsourcing contests have been used to develop health messages [7], inform health policy [8], and improve medical diagnostics [9]. These kinds of contests can increase community engagement [7,10], improve health [10,11], and save money [11].

However, crowdsourcing contests introduce a number of potential ethical concerns [12,13], including not being sufficiently inclusive, only relying on the internet, and not disseminating the solution widely. Identifying and responding to these shortcomings is important for establishing crowdsourcing as a force for the public good and as a useful public health tool. These concerns have received limited attention in the public health literature on crowdsourcing to date [2,4]. This paper describes crowdsourcing contests by stage, describes common ethical challenges, and provides guidance on implementing crowdsourcing contests ethically.

Methods

We conducted a scoping review [14] to synthesize literature on the ethical conduct of health-related crowdsourcing projects. This review includes applied and theoretical ethics literature related to crowdsourcing research and practice. Scoping reviews allow one to examine the literature in a structured way but are different from systematic reviews in their methodology and content [15]. Our review focused on sources between January 1, 2005 and July 1, 2017. We examined a wide range of anthropological, ethical, social science, and related literature on crowdsourcing to promote public health. We anchored this discussion in a particular example of a single crowdsourcing contest. In addition, we examined several crowdsourcing contest failures to understand concerns and potential ethical problems.

We identified studies using keyword searches in electronic databases, including MEDLINE (OVID interface, 1946 onwards), Google Scholar, expert opinion, and Wikipedia. For database searches, we used phrases and synonymous variations of the following terms: crowdsourcing, innovation challenge, ethics, implementation ethics, and applied ethical analysis. We also identified studies based on searches of reference lists, hand-searching key journals identified from initial database inquiries, and unpublished conference abstracts. We prioritized studies that examined crowdsourcing contests in health contexts. Our search included studies that provided empirical or theoretical data on crowdsourcing contests in the past 12 years.

Results

Overview

Our scoping review data are organized according to the 6 stages of a crowdsourcing contest—organizing, soliciting, promoting, judging, recognizing, and sharing [7,10]. First, the contest organizers form a contest steering committee to articulate the purpose, values, and methods of the contest. Second, an open call for content (eg, concepts, images, videos, or other materials) is announced via in-person events and social media. This open call clarifies the goals and terms of the contest, the prize or incentive structure, and the nature of participation. The open call plays a key role in defining the crowd. Third, the crowd is iteratively engaged through feedback sessions, in-person events, and social media. Fourth, a group of judges evaluates each contribution based on prespecified criteria to determine finalists. In some cases, the judges are the crowd itself. Finalists and others are awarded prizes according to their rank order. The judging process aggregates crowd wisdom [16]. Fifth, contest finalists are announced and recognized through an incentive structure. Sixth, the steering committee shares the finalist solution(s) with the community. After discussing each of these 6 stages, we review the literature on failures in crowdsourcing and discuss ethical principles of crowdsourcing contests.

Organizing a Steering Committee

The first step of a crowdsourcing contest is to establish a steering committee that will decide the structure and function of the contest. The steering committee powerfully shapes the contest and provides a set of norms, expectations, and deadlines. Often contests are divided into ones that focus on engaging large numbers of the community or on resulting in a high-quality outcome [4]. The condom video contest in China was focused on creating a high-quality video. The condom video contest steering committee was composed of youth, community health leaders, men who have sex with men, doctors, business leaders, and researchers. The group was organized by Sesh Global, an organization with experience in crowdsourcing contests. The steering committee met on a monthly basis to discuss the scope, rules, and promotion of the contest. In addition, the steering committee used email and social media to discuss contest developments.

One potential ethical problem with organizing a steering committee is the possibility of excluding community members or voices that are important to the contest. Often individuals from marginalized, vulnerable groups who lack a voice in decision making are less likely to be represented on steering committees. This problem has important implications for contests because the steering committee establishes the expectations and rules governing the entire process. For example, a contest focused on gay men and HIV ought to include gay men and people living with HIV. A committee lacking appropriate representation of key groups could undermine both effectiveness and trust in the contest.

One way to mitigate the risk of excluding important community voices is to have transparent criteria for selecting steering committee members. In addition, aligning the composition of the steering committee with the overall purpose of the contest could help ensure that community voices are represented. Given local power dynamics related to nonexpert advice, it may also be useful to have local, in-person meetings of the steering committee specifically to establish trust and align expectations.

Soliciting Contributions

Contest organizers design the open call soliciting contributions. The call for contributions is open so that anyone can contribute. Open calls can be through social media, in-person, or both. We define social media as websites or apps that allow users to create and share content or to engage in social networking [17]. Our condom video contest call for entries in China shows how the language, format, and structure shape a crowd (Multimedia Appendix 1). The call for entries was distributed through social media and in-person events at local high schools, colleges, and community-based organizations. The choice of distribution channels encouraged young Chinese individuals to participate but allowed entries from anyone.

One potential problem with calls for entries is over-reliance on social media announcements and insufficient attention to in-person events. Most private sector contests have focused on using social media calls to solicit entries [18], including several calls exclusively through social media [19,20]. There are two ethical problems with exclusively social media open calls. First, there is still a substantial digital divide between those who use social media and those who do not. Individuals who use social media tend to be from more developed regions or sectors and have higher socioeconomic status compared with those who do not use social media [21]. An exclusively social media call would not only constrain participation among some vulnerable groups, but it could worsen some of the entrenched social inequities. Second, among those who use social media, there is a further barrier to engaging sophisticated contest platforms, such as Ideascale, a company that creates online platforms for crowdsourcing contests [22]. A wide range of these platforms have been developed to crowdsource tasks. However, individuals who have the skills, knowledge, and experience to participate in these online platforms are a subset of the crowd, skewing its composition and unfairly excluding those without these skills but who are interested and could meaningfully contribute.

Careful attention to soliciting contributions can help to deal with these problems. In-person contest promotion events are one mechanism to broaden access to contests and diversify the crowd. These events have been used in several health contests [7] and have been found to increase participation and quality of participant entries. In-person events could take the form of classroom didactics, interactive feedback sessions, or community-led events. Capacity building sessions [23] could help individuals to learn about contributing on social media platforms. In addition to in-person contest promotion, ensuring multiple channels for contributing would be useful. This could include providing contributions through mail, in-person, or short text message. Contests should be as inclusive as possible relative to the intended audience. It is important to note that the goal is not for universal participation but to provide an opportunity to participate to those who would have a reasonable expectation of contributing to the contest.

Promoting Crowd Engagement and Contributions

Following the open call, there is an iterative process of engagement between organizers and potential participants. Our condom contest organized in-person and social media engagement activities to promote submissions. This included integration of in-person and social media activities so that they complemented each other. Approximately three-quarters of those who submitted to the contest participated in at least one engagement activity. These activities established trust in the contest, built confidence in contributing, and established social norms about how to participate in the contest. Engagement activities avoided giving examples in order to decrease cognitive fixation and increase innovation [24,25]. However, this stage of crowdsourcing contests also raises potential ethical concerns. Disclosure of confidential information by contributors and the possibility of social media trolling are two primary concerns.

Authentic engagement in a contest allows those who contribute to draw on their own unique talents, preferences, and local social context. However, this personal process introduces the risk of private information being divulged, often unintentionally, as part of engagement and contributing. These concerns have been raised more generally in the crowdsourcing literature [26]. In our case, some condom videos included identifiable individuals. The contest organizers had clear guidelines establishing that all videos could be publically viewed and any individual who participated in the video gave permission to be included. This decreased the risk of unintended disclosure associated with viewing the videos. In addition, contest organizers may consider having more stringent requirements about obtaining written consent for an individual's photograph or other personal information to be included in the contribution or going back to finalists to confirm consent before video or other forms of dissemination.

In addition, social media trolling has been reported within crowdsourcing contests. The word troll comes from the Scandinavian mythology, referring to evil small creatures who disturb travelers [17]. Today the term "trolling" refers to individuals who (usually anonymously) harass, provoke, or insult others online [27]. Trolling has been reported in a range of social media contexts [28,29], including contests [30,31]. Although trolling may be largely protected in some countries by the right to free speech [32], organizers of crowdsourcing

contests should make efforts to anticipate these harms and provide protection. This type of ethical concern can to some extent be addressed through online platform moderation and algorithms for detecting offensive words, as well as informing participants of potential risks of trolling during the consent process.

Aggregating Crowd Wisdom: Judging Contributions

Once the crowd has engaged in the contest and submitted contributions, these contributions are judged. The condom contest videos were evaluated by a multisectoral group of local people living with HIV, youth, physicians, and public health experts. Local judging increased community ownership of the contest and increased the likelihood that local characteristics (eg, using the local dialect) would be incorporated. At the same time, the nature of judging brings up a range of potential ethical issues, including how to fairly select judges. One common approach to judging contests has been to involve the crowd in judging entries, cited as a cost-effective way to engage potential participants [33]. However, if the crowd is exclusively defined online, it would be prone to the same problems described above in addition to bias, inconsistent judging criteria, and favoring popular opinion [34]. Crowd evaluation may also lead to voting based on criteria not consistent with the goals of the contest. For example, the British contest to name a government research vessel resulted in the entry "Boaty McBoatface" receiving 124,109 votes, more than fourfold greater than the next entry [35]. Organizers found themselves in the dilemma of accepting an absurd name or rejecting a crowdsourced outcome. They eventually compromised by using "Boaty McBoatface" to name a submersible carried by the research vessel dubbed the *Sir David Attenborough* [36]. In addition, the contest contributor with the largest number of online followers may be more likely to receive votes in support of their contest entry. Empirical evidence from private sector contests confirms that online crowd evaluation is biased toward individuals with greater social networks compared with expert judge evaluation [33]. Two studies found that individuals who win crowd-judged prizes are not as likely to sustain their engagement over time compared with individuals who win expert-judged prizes [33,37].

When a crowdsourcing contest has a relatively low number of entries (<100), a panel of expert judges could evaluate contributions. Judges would need to be selected in a fair way that is consistent with the mission and goals of the overall contest. Elements from "fair process" procedures—which emphasize transparency, justification of rationales, opportunities to appeal decisions, and so on—could help in the constitution of the judging panel and also help guide the decisions they make [38]. This could include evaluation of the judge panel to help ensure that a broad range of judges are represented and decrease reliance on social media. Several private sector contests demonstrate the feasibility of having a judging panel evaluate contributions [39].

Celebrating Crowd Wisdom: Recognizing Contributions

Following the judging process, contributions can be recognized by prizes or incentives, acknowledgment, and retention of legal rights to products created. Incentive structures for crowdsourcing vary based on the goals and missions of the contest. Some contests have a single large prize [40] while others recognize a number of contributions [41]. The condom contest included individual prizes for the top three contributors as well as participation prizes. The top contributors were announced on social media as a further form of recognition. All contributors retained the rights to their videos (those who submitted the videos could use them for any purposes), consistent with the goal of the contest to promote community agency.

An ethical challenge related to recognition in crowdsourcing contests is the potential for exploiting those who make substantial contributions. Insufficient recognition of those who contribute to contests has been noted in many online contest settings [42-44]. The condom contest decreased the likelihood of this exploitation because there were several formal and informal ways of recognizing participants, alongside retention of their legal rights. The contest also shared the finalist video online in several forums.

Appropriately recognizing contributions provides a way of addressing these concerns about crowd exploitation related to the incentive structure and acknowledgment. First, clearly stating during the consent process how contributions will be recognized can mitigate exploitation to some extent. Second, incentive structures with multiple prizes (of different types) promote a broad spectrum of participation. Including special prize categories that focus on participation rather than merit have been used in some contests [7]. Third, formally acknowledging and celebrating contributions are important. Several studies have shown that intrinsic benefits of participation (such as recognition and media attention) are more important than extrinsic benefits in the context of crowdsourcing [10]. Governance of ownership and permissible uses of finalist contributions may also minimize exploitation.

Sharing and Implementing the Solution With the Community

The final stage of a crowdsourcing contest is to share the solution more widely with the community that contributed. Henk van Ess has argued that crowdsourcing must give back to the public and share the solution more widely [45]. In this way, crowdsourcing reciprocates in a commensurate way to what the community contributed. Other crowdsourced research has suggested that perceptions of fairness are important for those contributing to crowdsourcing projects [46,47]. Our condom video contest provided prizes to finalists, participation prizes, and then made the videos available on public platforms in China.

Limited sharing of the crowdsourced solution presents an important ethical concern associated with crowdsourcing contests. This also differentiates public-oriented contests from their private sector counterparts. Most private sector contests see the finalist solutions as their own intellectual property; the terms of many private contests give intellectual property rights to organizers. Limited sharing could take the form of only describing contests in articles that are inaccessible to nonsubscribers behind a paywall.

Clearly establishing a plan early in the process for sharing and prizes as part of the call for entries can mitigate the risk of

insufficient sharing. While having clear sharing expectations is important, there should also be sufficient flexibility to give the steering committee ultimate authority in making final decisions. For example, the condom contest mandated sharing of the final video on regional, national, and international networks. The call for entries specified this plan, in addition to a plan to recognize excellent entries. The benchmark for "excellent" was decided by the steering committee.

Learning From Failures

Previous examples of crowdsourcing contests that were partially or incompletely effective provide guidance (Table 1). In 2013, the condom manufacturing company Durex invited the public to vote on which city in the world should receive a special condom rush delivery service [48]. By the end of the contest, the nonexistent city of "Batman" in Turkey had received the most votes, and contest organizers were faced with the ethical dilemma of blatantly rejecting the clear will of the crowd or endorsing a deliberately facetious winning entry. Such instances of crowd hijacking are not uncommon [49], and contest organizers should prepare for scenarios where the crowd may use the contest platform to advance an agenda that deviates from that of the contest organizers'. Given the unpredictability of the crowd, it is important for organizers to clearly explain their rights to prospective participants, including the right to deem certain kinds of entries inadmissible.

Past contests have also shown that successfully developing products through crowdsourcing is not formulaic, and that breakthrough innovations are by no means guaranteed. The start-up company Quirky had managed to secure hundreds of millions of investment dollars to develop innovative household consumer products through open online contests [50]. However, despite deep financial resources and hundreds of thousands of contest contributors, Quirky failed to produce any radically innovative products and eventually declared bankruptcy 6 years after its founding. One of the major problems was that Quirky innovators had disagreements with the company in the late stages of business development [50]. This miscommunication could be avoided by involving community members earlier in the process of development.

Finally, a German contest solicited public input on a ban on circumcision. A political party decided to crowdsource local opinions on the topic, targeting the area of North Rhine-Westphalia. Despite having a population of 18 million individuals, the contest only received 20 submissions [51]. This underscores the importance of having a steering committee that plans in advance and understands community interests and willingness to take part in crowdsourcing contests.

Ethical Principles in Crowdsourcing Contests

We identified several general reviews that broadly considered ethical principles associated with crowdsourcing contests [12,13,52]. These highlighted theoretical concerns about privacy, accuracy of information, property, and accessibility in the context of computer science. However, this limited literature did not focus on health contests.

Table 1. Implementation ethics issues and potential solutions associated with crowdsourcing contests.

Contest stages	Implementation ethics issue	Potential solution
1 Organizing	Lack of input from community voices or marginalized groups	Explicitly state criteria for selecting steering committee members to ensure adequate representation
2 Soliciting	Online contests limit participation to a subset of internet-using individuals	In-person events to promote contests; multiple ways of receiving contributions
	Social networking sites narrow participation in contests to a subset of social media-savvy individuals	Allow contributions via email, in-person, cell phones, and other forms that do not require online access or social media
3 Promoting	Public contributions may include confidential or private information	Clear contest guidelines that clarify whose permission has been obtained and potentially enhanced consent process before dissemination
	Social media platforms for contributing may introduce opportunities for online harassment	Social media moderators and algorithms for detection of explicit language
4 Judging	Crowd evaluation may be biased in favor of online individuals with larger social networks	Form a local judge panel composed of key individuals representing different perspectives or backgrounds
	Multiple ways of selecting judges	Establish guidelines for selecting judges and transparent procedures for evaluation and judging
5 Recognizing	Single prize contests that are most optimal provide no recognition for most contributors	Multiple prize or incentive structure encourages a broad range of participation
	Online contests may not sufficiently recognize contributions	In-person prize announcements
6 Sharing	More is taken from the community than given back	Establish a formal mechanism to share or implement the solution more widely with the local community

Discussion

Principal Findings

Our review suggests that crowdsourcing contests for public health introduce several potential ethical concerns. These concerns can be categorized within the 6 stages of crowdsourcing contests—organizing, soliciting, promoting, judging, recognizing, and sharing. Our analysis suggests that these concerns can be minimized with appropriate planning and consultation. Our review expands the limited literature on crowdsourcing contests for public health [4] by focusing on ethics, examining an empirical case, and including a formal scoping review.

Our data suggest that several ethical concerns associated with crowdsourcing contests can be anticipated and avoided. For example, the awkward situation of having crowds decide on a trivial name such as *Boaty McBoatface* can be avoided by separating the process of soliciting names and choosing names eligible to be voted on. Other risks can be mitigated through appropriate contest planning. For example, inviting a diversity of local steering group members can increase the likelihood of local community perspective representation. Other ethical concerns are related to the use of social media within contests (eg, privacy concerns) and have been well described elsewhere [53]. All of these types of ethical concerns underscore the need for organizers of crowdsourcing contests to include sufficient time for planning and designing a contest.

Our scoping review did not identify studies that articulated ethical principles of crowdsourcing contests for health. This may be because few studies have focused on using crowdsourcing contests to improve public health [4]. This also may be related to the breadth of diversity of crowdsourcing contests activities, including research studies, community engagement programs, and communications strategies. However, our analysis suggests that there are several shared contest stages each of which has potential ethical concerns.

Crowdsourcing contests have implications for public health research and policy. In terms of research, further empirical study on ethical problems associated with crowdsourcing contests is necessary. Such research could help to refine the method and increase the likelihood of crowdsourcing contests achieving their goals. This research could include qualitative studies of those participating and organizing crowdsourcing contests [54]. Such research would provide valuable input for a future ethical framework specific to crowdsourcing. In terms of policy, crowdsourcing contests could help to inform public health policy. The multisectoral, transparent, and open nature of contests establishes a strong foundation for policy making. For example, the World Health Organization (WHO) used a crowdsourcing contest [55] to solicit descriptions of hepatitis testing that were directly included in the 2017 WHO hepatitis testing guidelines [56]. Given the potential for contests to inform policy, more formal principles and ethical considerations associated with crowdsourcing contests may be useful.

Limitations

Our analysis has several limitations. First, most health contests have been single events and have yet to be serialized and formally incorporated into routine public health practice. Serial contests are likely to have different ethical challenges and may be substantially different in terms of implementation. Second, health-focused crowdsourcing contests are relatively new and while there are many examples of health-related contests, few are formally evaluated using validated metrics. Our analysis focused on a single example. Further implementation research is needed to define the most efficient and responsible use of crowdsourcing contests. Third, there is variation in the extent to which contests are driven by crowd input. Some health projects involve the community at all stages [57], while others have more intensive community input only at the start of the project [4].

Conclusions

Crowdsourcing contests may be a useful tool to develop inclusive public health programs but also pose ethical concerns at each stage. Unraveling these ethical concerns requires careful planning, consideration, and consultation. Our analysis provides several practical steps for the responsible conduct of crowdsourcing contests and identifies areas for future research.

Acknowledgments

The authors express their gratitude to Weiming Tang, Songyuan Tang, and Eric Juengst for helpful comments on a previous version of this manuscript. The condom contest research described in this methodology article was approved by the Guangdong Provincial STD Control Institutional Review Board (IRB; 114310-01) and the University of North Carolina Chapel Hill IRB (Study 15-1522). All individuals completed informed consent. All research data are available upon request, contingent upon the approval of Chinese and American IRBs. This research was supported by research grants from the National Institute of Allergy and Infectious Diseases (1R01AI114310, 1R01A108366).

Authors' Contributions

JT wrote the first draft of this manuscript. SP and GT revised the section on crowdsourcing failures. BB and AM provided guidance on writing on the crowdsourcing evaluation part. SR contributed to the implementation ethics sections. All authors contributed to the writing of the manuscript and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Call for entries from the crowdsourcing contest.

[PDF File (Adobe PDF File), 38KB - jmir_v20i3e75_app1.pdf]

References

1. Oxford English Dictionary. Discover the story of English URL: http://www.oed.com/ [accessed 2018-02-15] [WebCite Cache ID 6xFHS4js9]
2. Brabham DC, Ribisl KM, Kirchner TR, Bernhardt JM. Crowdsourcing applications for public health. Am J Prev Med 2014 Feb;46(2):179-187. [doi: 10.1016/j.amepre.2013.10.016] [Medline: 24439353]
3. Wikipedia. URL: https://en.wikipedia.org/wiki/Wikipedia [accessed 2017-06-11] [WebCite Cache ID 6r8u98mVw]
4. Pan S, Stein G, Bayus B, Tang W, Mathews A, Wang C, et al. Systematic review of innovation design contests for health: spurring innovation and mass engagement. BMJ Innov 2017 Sep 27;3(4):227-237. [doi: 10.1136/bmjinnov-2017-000203]
5. Cooper S, Khatib F, Treuille A, Barbero J, Lee J, Beenen M, et al. Predicting protein structures with a multiplayer online game. Nature 2010 Aug 05;466(7307):756-760 [FREE Full text] [doi: 10.1038/nature09304] [Medline: 20686574]
6. Ringh M, Rosenqvist M, Hollenberg J, Jonsson M, Fredman D, Nordberg P, et al. Mobile-phone dispatch of laypersons for CPR in out-of-hospital cardiac arrest. N Engl J Med 2015 Jun 11;372(24):2316-2325. [doi: 10.1056/NEJMoa1406038] [Medline: 26061836]
7. Zhang Y, Kim JA, Liu F, Tso LS, Tang W, Wei C, et al. Creative contributory contests to spur innovation in sexual health: 2 cases and a guide for implementation. Sex Transm Dis 2015 Nov;42(11):625-628 [FREE Full text] [doi: 10.1097/OLQ.0000000000000349] [Medline: 26462186]
8. Hildebrand M, Ahumada C, Watson S. CrowdOutAIDS: crowdsourcing youth perspectives for action. Reprod Health Matters 2013 May;21(41):57-68. [doi: 10.1016/S0968-8080(13)41687-7] [Medline: 23684188]
9. Wolf M, Krause J, Carney PA, Bogart A, Kurvers RH. Collective intelligence meets medical decision-making: the collective outperforms the best radiologist. PLoS One 2015;10(8):e0134269 [FREE Full text] [doi: 10.1371/journal.pone.0134269] [Medline: 26267331]
10. Merchant RM, Griffis HM, Ha YP, Kilaru AS, Sellers AM, Hershey JC, et al. Hidden in plain sight: a crowdsourced public art contest to make automated external defibrillators more visible. Am J Public Health 2014 Dec;104(12):2306-2312. [doi: 10.2105/AJPH.2014.302211] [Medline: 25320902]
11. Tang W, Han L, Best J, Zhang Y, Mollan K, Kim J, et al. Crowdsourcing HIV testing: a pragmatic, non-inferiority randomized controlled trial in China. 2015 Presented at: IAS 2015; July, 2015; Vancouver. [doi: 10.1093/cid/ciw171]
12. Alqahtani BA, El-shoubaki RT, Noorwali FA, Allouh D, Hemalatha M. 2017. Legal and ethical issues of crowdsourcing URL: http://www.ijcaonline.org/archives/volume167/number10/alqahtani-2017-ijca-914324.pdf [accessed 2018-01-09] [WebCite Cache ID 6wL2nUIUu]
13. Schmidt FA. 2013. The good, the bad and the ugly URL: http://florianalexanderschmidt.de/the-good-the-bad-and-the-ugly/ [accessed 2018-01-09] [WebCite Cache ID 6wL2ybNd6]
14. 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]
15. Mays N, Roberts E, Popay J. Synthesising research evidence. In: Fulop N, Allen P, Clarke A, Black N, editors. Studying the organisation and delivery of health services: research methods. London: Routledge; 2001:188-220.
16. Surowiecki J. The wisdom of crowds : why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations. New York: Doubleday; 2004.
17. Stevenson A, editor. Oxford dictionary of English (3 ed). Oxford, England: Oxford University Press; 2014.
18. Bullinger AC, Moeslein K. 2013 Apr. Innovation Contests - Where are we? URL: http://aisel.aisnet.org/cgi/viewcontent.cgi?article=1027&context=amcis2010
19. Carpenter J. May the best analyst win. Science 2011 Feb 11;331(6018):698-699. [doi: 10.1126/science.331.6018.698] [Medline: 21311005]
20. Hussein A. Brokering knowledge in biosciences with InnoCentive. Interview by Semahat S. Demir. IEEE Eng Med Biol Mag 2003;22(4):26-27. [Medline: 14515687]
21. Norris P. Digital Divide: Civic Engagement, Information Poverty, and the Internet Worldwide. New York: Cambridge University Press; 2001.
22. Ideascale. URL: https://ideascale.com/ [accessed 2017-06-12] [WebCite Cache ID 6rAnsk4SR]
23. Shapira N, Barak A, Gal I. Promoting older adults' well-being through internet training and use. Aging Ment Health 2007 Sep;11(5):477-484. [doi: 10.1080/13607860601086546] [Medline: 17882585]
24. Smith SM, Ward TB, Schumacher JS. Constraining effects of examples in a creative generation task. Mem Cognit 1993 Nov;21(6):837-845. [Medline: 8289661]
25. Marsh RL, Landau JD, Hicks JL. How examples may (and may not) constrain creativity. Mem Cognit 1996 Sep;24(5):669-680. [Medline: 8870535]

26. Durward D, Blohm I. Is there PAPA in crowd work? A literature review on ethical dimensions in crowdsourcing. 2016 Presented at: IEEE International Conference on Internet of People; 2016; Tolouse, France. [doi: 10.1109/UIC-ATC-ScalCom-CBDCom-IoP-SmartWorld.2016.0131]
27. Bartow A. Internet defamation as profit center: the monetization of online harassment. Harv JL & Gender 2009;32(2):1-48 [FREE Full text]
28. Shin J. Morality and internet behavior: a study of the internet troll and its relation with morality on the internet. Chesapeake, VA: Association for the Advancement of Computing in Education; 2008 Presented at: Society for Information Technology & Teacher Education International Conference; 2008; Nevada, Las Vegas.
29. Bishop J. Representations of 'trolls' in mass media communication: a review of media-texts and moral panics relating to 'internet trolling'. IJWBC 2014;10(1):7-24. [doi: 10.1504/IJWBC.2014.058384]
30. Freischlad N. Jovoto. 2018. Something about trolls URL: http://www.jovoto.com/blog/2010/03/something-about-trolls/ [WebCite Cache ID 6r8ubUekf]
31. Ekman O. Jovoto. 2018. To rate or not to rate URL: http://www.jovoto.com/blog/2009/12/to-rate-or-not-to-rate/ [WebCite Cache ID 6r8ueRLEH]
32. Diaz FL. 2016. Trolling and the first amendment: protecting internet speech in an era of cyberbullies and internet defamation URL: http://illinoisjltp.com/journal/wp-content/uploads/2016/06/Diaz.pd [accessed 2018-02-16] [WebCite Cache ID 6xGf8oCWd]
33. Chen L, Xu P, Liu D. New York: Elsevier; 2016. Experts versus the crowd: a comparison of selection mechanisms in crowdsourcing contests URL: https://tinyurl.com/yamftjow
34. Aral S. 2013 Dec 19. The Problem with online ratings URL: https://sloanreview.mit.edu/article/the-problem-with-online-ratings-2/
35. Ellis-Petersen H. The Guardian. 2016. Boaty McBoatface wins poll to name polar research vessel URL: https://www.theguardian.com/environment/2016/apr/17/boaty-mcboatface-wins-poll-to-name-polar-research-vessel [accessed 2018-02-20] [WebCite Cache ID 6xMhNEVEA]
36. Knapton S. The Daily Telegraph. London; 2016. 'BoatyMcBoatface' to live on as yellow submarine, science minister Jo Johnson announces URL: http://www.telegraph.co.uk/science/2016/05/06/boatymcboatface-to-live-on-as-yellow-submarine-science-minister/ [accessed 2018-02-20] [WebCite Cache ID 6xMhUg0TA]
37. Yang J, Adamic L. Crowdsourcing and knowledge sharing: strategic user behavior on taskscn. 2008 Presented at: 9th ACM Conference on Electronic Commerce; 2018; Chicago, USA p. 246-255. [doi: 10.1145/1386790.1386829]
38. Daniels N. Just Health Care. Cambridge, UK: Cambridge University Press; 1985.
39. Issue Lab. And the winner is...: capturing the promise of philanthropic prizes URL: https://www.issuelab.org/resource/and-the-winner-is-capturing-the-promise-of-philanthropic-prizes.html [accessed 2017-06-11] [WebCite Cache ID 6r8ukWx0w]
40. Moldovanu B, Sela A. The optimal allocation of prizes in contests. Am Econ Rev 2001;91(3):542-548. [doi: 10.1257/aer.91.3.542]
41. Sisak D. Multiple-prize contests - the optimal allocation of prizes. J Econ Surv 2009;23(1):82-114. [doi: 10.1111/j.1467-6419.2008.00557.x]
42. Ross J, Irani L, Silberman S, Zaldivar A, Tomlinson B. Who are the crowdworkers? Shifting demographics in mechanical Turk. In: ACM CHI. 2010 Presented at: 28th International Conference on Human Factors in Computing Systems; 2010; Atlanta, Georgia. [doi: 10.1145/1753846.1753873]
43. Kittur A, Nickerson J, Bernstein M, Gerber E, Shaw A, Zimmerman J, et al. The future of crowd work. In: Proceedings of the 2013 conference on Computer supported cooperative work. 2013 Presented at: CSCW '13; 2013; San Antonio. [doi: 10.1145/2441776.2441923]
44. Graber MA, Graber A. Internet-based crowdsourcing and research ethics: the case for IRB review. J Med Ethics 2013 Feb;39(2):115-118. [doi: 10.1136/medethics-2012-100798] [Medline: 23204319]
45. van Ess H. SlideShare. 2010. Crowdsourcing: How to find a crowd URL: https://www.slideshare.net/searchbistro/harvesting-knowledge-how-to-crowdsource-in-2010 [accessed 2018-02-15] [WebCite Cache ID 6xFJxJSl6]
46. Faullant R, Fueller J, Hutter K. Fair play: perceived fairness in crowdsourcing communities and its behavioral consequences. Acad Manage Proc 2013 Nov 25;2013(1):15433. [doi: 10.5465/AMBPP.2013.15433abstract]
47. Mazzola E, Piazza M, Acur N. The Impact of Fairness on the Performance of Crowdsourcing: An Empirical Analysis of Two Intermediate Crowdsourcing Platforms. 2016 Presented at: EURAM Conference Proceedings; 1-4 June, 2016; Paris, France.
48. Lutz A. Business Insider. New York: Business Insider; 2013. How Durex's Social Media Contest For Condoms Totally Backfired URL: https://www.businessinsider.in/how-durexs-contest-totally-backfired-2013-6?r=US&IR=T
49. Wilson M, Robson K, Botha E. Crowdsourcing in a time of empowered stakeholders: lessons from crowdsourcing campaigns. Bus Horiz 2017;60(2):247-253. [doi: 10.1016/j.bushor.2016.11.009]
50. Fixson S, Marion TJ. 2016 Dec 15. A case study of crowdsourcing gone wrong URL: https://hbr.org/2016/12/a-case-study-of-crowdsourcing-gone-wrong [accessed 2018-02-16] [WebCite Cache ID 6xGtKZ78Y]

51. Dahlander L, Piezunka H. 2017. Why some crowdsourcing efforts work and others don't URL: https://hbr.org/2017/02/why-some-crowdsourcing-efforts-work-and-others-dont [accessed 2018-02-20] [WebCite Cache ID 6xMiBJnIb]
52. Standing S, Standing C. The ethical use of crowdsourcing. Bus Ethics 2017 Oct 10;27(1):72-80. [doi: 10.1111/beer.12173]
53. Golder S, Ahmed S, Norman G, Booth A. Attitudes toward the ethics of research using social media: a systematic review. J Med Internet Res 2017 Jun 06;19(6):e195 [FREE Full text] [doi: 10.2196/jmir.7082] [Medline: 28588006]
54. Sugarman J, Sulmasy DP. Methods in Medical Ethics, 2nd Ed. Washington DC, US: Georgetown University Press; 2010.
55. Tucker JD, Meyers K, Best J, Kaplan K, Pendse R, Fenton KA, et al. The HepTestContest: a global innovation contest to identify approaches to hepatitis B and C testing. BMC Infect Dis 2017 Nov 01;17(Suppl 1):701 [FREE Full text] [doi: 10.1186/s12879-017-2771-4] [Medline: 29143673]
56. World Health Organization. Geneva; 2017. WHO guidelines on hepatitis B and C testing URL: http://apps.who.int/iris/bitstream/10665/254621/1/9789241549981-eng.pdf [accessed 2018-02-20] [WebCite Cache ID 6xMiJBd75]
57. SESH Study Group, Tucker JD. Crowdsourcing to promote HIV testing among MSM in China: study protocol for a stepped wedge randomized controlled trial. 2017 Oct 02;18(1):447 [FREE Full text] [doi: 10.1186/s13063-017-2183-1] [Medline: 28969702]

Abbreviations

IRB: Institutional Review Board
WHO: World Health Organization

Edited by G Eysenbach; submitted 16.06.17; peer-reviewed by M Graber, WH Sun, J Jiang; comments to author 22.11.17; revised version received 09.01.18; accepted 09.01.18; published 09.03.18.

Please cite as:
Tucker JD, Pan SW, Mathews A, Stein G, Bayus B, Rennie S
Ethical Concerns of and Risk Mitigation Strategies for Crowdsourcing Contests and Innovation Challenges: Scoping Review
J Med Internet Res 2018;20(3):e75
URL: http://www.jmir.org/2018/3/e75/
doi:10.2196/jmir.8226
PMID:29523500

Original Paper

Investigating the Extent to Which Patients Should Control Access to Patient Records for Research: A Deliberative Process Using Citizens' Juries

Mary P Tully[1,2], BSc (Hons), MSc, PhD, FRPharmS, FFRPS; Kyle Bozentko[3], MTS; Sarah Clement[4], BSc, PhD; Amanda Hunn[5], BSc (Hons), MA; Lamiece Hassan[1], BSc (Hons), MPhil, PhD; Ruth Norris[1], BSc (Hons), PGDip; Malcolm Oswald[6,7], BA (Econ), MA, PhD; Niels Peek[1,8], PhD

[1]Health E-Research Centre, Division of Imaging, Informatics and Data Sciences, School of Health Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester Academic Health Science Centre, Manchester, United Kingdom

[2]Division of Pharmacy and Optometry, School of Health Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester Academic Health Science Centre, Manchester, United Kingdom

[3]Jefferson Center, Saint Paul, MN, United States

[4]Department of Geography and Planning, School of Environmental Sciences, University of Liverpool, Liverpool, United Kingdom

[5]Health Research Authority, London, United Kingdom

[6]School of Law, Faculty of Humanities, The University of Manchester, Manchester, United Kingdom

[7]Citizens Juries Community Interest Company, Manchester, United Kingdom

[8]Greater Manchester Patient Safety Translational Research Centre, Division of Population Health, Health Services Research and Primary Care, School of Health Sciences, The University of Manchester, Manchester Academic Health Science Centre, Manchester, United Kingdom

Corresponding Author:
Mary P Tully, BSc (Hons), MSc, PhD, FRPharmS, FFRPS
Division of Pharmacy and Optometry
School of Health Sciences, Faculty of Biology, Medicine and Health
The University of Manchester, Manchester Academic Health Science Centre
Oxford Road
Manchester, M13 9PL
United Kingdom
Phone: 44 01612754242
Email: mary.p.tully@manchester.ac.uk

Abstract

Background: The secondary use of health data for research raises complex questions of privacy and governance. Such questions are ill-suited to opinion polling where citizens must choose quickly between multiple-choice answers based on little information.

Objective: The aim of this project was to extend knowledge about what control *informed* citizens would seek over the use of health records for research after participating in a deliberative process using citizens' juries.

Methods: Two 3-day citizens' juries, of 17 citizens each, were convened to reflect UK national demographics from 355 eligible applicants. Each jury addressed the mission "To what extent should patients control access to patient records for secondary use?" Jurors heard from and questioned 5 expert witnesses (chosen either to inform the jury, or to argue for and against the secondary use of data), interspersed with structured opportunities to deliberate among themselves, including discussion and role-play. Jurors voted on a series of questions associated with the jury mission, giving their rationale. Individual views were polled using questionnaires at the beginning and at end of the process.

Results: At the end of the process, 33 out of 34 jurors voted in support of the secondary use of data for research, with 24 wanting individuals to be able to opt out, 6 favoring opt in, and 3 voting that all records should be available without any consent process. When considering who should get access to data, both juries had very similar rationales. Both thought that public benefit was a key justification for access. Jury 1 was more strongly supportive of sharing patient records for public benefit, whereas jury 2 was more cautious and sought to give patients more control. Many jurors changed their opinion about who should get access to health records: 17 people became more willing to support wider information sharing of health data for public benefit, whereas 2 moved toward more patient control over patient records.

Conclusions: The findings highlight that, when informed of both risks and opportunities associated with data sharing, citizens believe an individual's right to privacy should not prevent research that can benefit the general public. The juries also concluded that patients should be notified of any such scheme and have the right to opt out if they so choose. Many jurors changed their minds about this complex policy question when they became more informed. Many, but not all, jurors became less skeptical about health data sharing, as they became better informed of its benefits and risks.

(J Med Internet Res 2018;20(3):e112) doi:10.2196/jmir.7763

KEYWORDS

public participation; patient engagement; public opinion; medical research; confidentiality; privacy; national health services; data linkage; public policy, decision making, organizational

Introduction

Public Opinion on Data Use

The last decade has seen a surge in the reuse of data that were created for the health care of individual patients for additional purposes such as for research (so-called secondary use of data). Research studies that previously would have been impossible because of the effort required to collect data have now become possible by reusing data originally collected for the purposes of providing direct health care. Examples include investigations into the prescribing of anxiolytics and hypnotics in over 300,000 children and young people in Wales [1], the mental health of 57,000 veterans compared with173,000 nonveterans in Scotland [2], and the impact of a smoke-free legislation on stroke [3].

In many countries, there is no lawful impediment to the use of deidentified (or anonymized) data for research without the consent of the data subject, as long as the risks of reidentification are very low or remote. This may include the linkage of data from multiple sources before deidentification. However, public support for such research use of data without consent, the so-called social license, is separate from any legal framework [4]. Even where no legislation exists to prevent the use of deidentified data, the lack of a social license may ultimately result in the failure of data-use initiatives, as has been the case with national data records systems in England [4] and Australia [5].

Epidemiologists are dependent upon using data without consent for such research for numerous reasons. Obtaining consent from many thousands of people is an onerous task, and there is a strong likelihood of many people being disinterested and not giving consent simply because they do not remember being asked [6]. Large amounts of missing data, which are often from particular subgroups rather than randomly distributed through the population, can mean that findings from epidemiological studies can be misleading [7].

Despite such opinions in the aggregate, individual public attitudes toward the secondary use of data vary [8,9]. In particular, public support may be different, depending on who is using the data or the use to which the data may be put. Such reuse of data without consent is an area of concern to some members of the public [10]. These people express the wish to be asked to consent to every use of the data, whereas others want to give a general consent for data use. Still others are content with the data being used without them being aware, consulted, or asked at all, provided that the research has been reviewed and approved by an ethics committee [9].

Deliberative Approaches

It could be argued that, for such a complex area, surveys may not be the best method to find out about the decisions that the public would make. Members of the public are often unaware of the ways that data are used and the governance procedures that are put in place to protect health care data [8,10,11]. Thus, it could be surmised that some survey respondents do so from a position of ignorance of the topic. Qualitative methods such as focus groups or deliberative processes provide a more nuanced view of public opinion. Some of these methods, particularly those which employ deliberative approaches, enable questions to be answered about what citizens would think regarding the use of data if they were informed. Few studies have done this [11,12]. Recently, the Wellcome Trust compared findings from a survey about the commercial use of data conducted with 2017 members of the public, with findings from 16 focus groups with 246 people [11]. The focus groups indicated how people change their minds once they were slightly more informed about the use of data through discussion with their peers in the group, whereas surveys are generally conducted at a single point in time.

This suggests that there is much to be learned from using deliberative methods that allow participants to learn about and reflect on information about such a complex area. Citizens' juries are comprehensive engagement processes that allow decision makers and the public to hear thoughtful input from an informed microcosm of the public [13]. They are based on the premise that, given enough time, opportunity, support, and resources, members of the public are quite capable of arriving at decisions about complex matters [13,14]. The citizens' jury process is designed to allow decision makers to hear citizens' voices. It provides an opportunity for citizens to learn about an issue and deliberate together to find a common ground solution. Decision makers can thus learn more about what an informed public wants and why they want it [13].

There are examples of organizations using citizens' juries to help make policy decisions, even though members of juries are not elected and cannot be made accountable for decisions. For example, Melbourne City Council has appointed a citizens' jury to determine how to allocate its 10-year Aus $5 billion budget, and the council is implementing virtually all of the jury's recommendations [15].

Study Aim

The aim of this study was to investigate what people think about secondary use of data, including data linkage, once they become more informed about the area. The outcome of the citizens' juries was to inform the ongoing research, information governance, and public engagement strategies of the project's sponsors: the UK's national Farr Institute of Health Informatics Research [16] and the Greater Manchester Primary Care Patient Safety Translational Research Centre [17].

Methods

Jury Process

The citizens' juries were run over 3 days as jury pairs [18], that is, two juries were conducted in the same geographical area, addressed the same jury mission (Textbox 1), listened to the same witnesses, but were comprised of different people. To ensure that the juries were conducted appropriately, the manual written by the Jefferson Center, the developers of the method, was followed [13], and the juries were run by an experienced facilitator from that center (KB). Approval for video recording of the jury discussions was obtained from The University of Manchester's research ethics committee.

Jury Recruitment

The citizens' jury process uses members who are selected to be representative of the population in key criteria [13]. Recruitment questionnaires collected data to enable selection against *a priori* criteria based on demographics and views on privacy (Table 1). The demographics provided a broadly representative sample of resident adults in England based on the 2011 census with respect to gender, age range, ethnicity, and educational attainment [19]. Potential participants were asked to complete an Ipsos MORI survey question [20] that involved balancing privacy against information sharing for public benefit (Textbox 2). This was the most up-to-date survey of public opinion in this area at the time of recruitment. The demographic and privacy criteria insured that each jury was a "microcosm of the public" [13].

Textbox 1. The jury mission.

Suppose a National Health Service (NHS) body wants to create new records from the patient records stored by your general practice and by hospitals that have treated you. They want to use them for purposes other than your direct patient care, such as research about better treatments and for checking that patients are receiving safe and effective health care. These records would be held securely and would not contain your name, address, and other identifiers. Despite this, there is a small risk that the records might still identify you because they would contain lots of detailed information about the care you receive from your general practitioner and from different hospitals. The NHS body would also review requests from other public and private organizations, granting access only where they believed it was lawful and in a good cause.

1. i. Should the NHS body be allowed to create these records about you and other patients? (choose only one of the following)
   - a. Yes, but they should publish information about what they plan to do
   - b. Yes, but they should publish information about what they plan to do and patients should be able to opt out
   - c. Yes, but they should publish information about what they plan to do and only create records for patients who opt in
   - d. No
   - e. Other (explain in less than 30 words)

   ii. Give reasons for your answer (in less than 300 words)

2. i. Given your answer to question 1, who should be allowed to access and extract data from the records created? (Choose as many of the following examples that apply)
   - a. NHS clinicians and administrators who decide which health services should (and should not) be funded
   - b. NHS clinicians and administrators doing approved research into whether doctors are prescribing medicines appropriately
   - c. University staff doing approved research into whether doctors are prescribing medicines appropriately
   - d. Staff employed by local authorities planning the future need for residential care homes
   - e. Staff employed by a private company being paid by a hospital NHS trust to compare the number of people dying after surgery with other hospitals
   - f. Staff employed by an insurance company aiming to set health insurance premiums accurately
   - g. Staff employed by a pharmaceutical company investigating whether they should begin research into a new drug for a genetic disease for which there is currently no treatment

   ii. Give reasons for your answer (in less than 400 words)

Table 1. A priori criteria for jury selection and demographics of actual jurors.

Criteria	UK census (%)[a]	Jury target range	Achieved in jury 1 and 2[b]
Gender			
Women	51	8-10 jurors	8 and 9 jurors
Men	49	8-10 jurors	9 and 8 jurors
Age range (years)			
18-29	21	2-5 jurors	5 and 3 jurors
30-44	26	3-6 jurors	4 and 6 jurors
45-59	25	3-6 jurors	5 and 5 jurors
60+	28	4-7 jurors	3 and 3 jurors
Ethnicity			
White	85	14-17 jurors	14 and 14 jurors
Groups other than white	15	2-4 jurors	3 and 3 jurors
Educational attainment			
Level 1 or no qualifications	36	5-8 jurors	6 and 7 jurors
Level 2 or level 3 qualifications (apprenticeship and other qualifications)	37	5-8 jurors	6 and 5 jurors
Level 4 qualifications (degree level) and above	27	4-6 jurors	5 and 5 jurors
Privacy views[c]			
Agree more with a) than b)	52	7-11 jurors	9 and 10 jurors
Agree more with b) than with a)	34	5-7 jurors	5 and 6 jurors
Agree equally with both or don't agree with either or don't know	14	1-4 jurors	3 and 1 jurors

[a][19].

[b]1 person left each jury at the end of the first day and are not reported here.

[c]Target sample percentages based on "Perceptions of Data Sharing" survey [20]—see Textbox 2 for full text.

Textbox 2. Ipsos MORI survey question used to assess views and privacy for jury selection and after the jury was completed.

As you may know, different government departments and services collect data about individuals, for example, your tax records and health records. People have different views on how much of this information should be shared within government. Data sharing can bring benefits such as finding more effective medical treatments, using information about local communities to plan local schools or roads, etc. But some people worry that data sharing will be a risk to their privacy and security, by linking different types of data together and potentially allowing them to be identified. Overall, which of the following statements is closest to your view?

a. We should share all the data we can because it benefits the services and me—as long as I can opt out if I choose

b. We should not share data as the risks to people's privacy and security outweigh the benefits

1. Agree much more with a) than with b)
2. Agree a little more with a) than with b)
3. Agree equally or don't agree or don't know
4. Agree a little more with b) than with a)
5. Agree much more with b) than with a)

Jury members were recruited using a variety of methods to ensure that the criteria were met. Adverts were placed on websites for employment opportunities and research volunteers, emails were sent to a range of community groups, and in-person presentations were made to groups of retired people. Most of the members were recruited from the employment website. From a jury pool of 355 eligible applicants, 18 jurors and 4 reserves were selected for each jury, as recommended by the Jefferson Center [13]. Candidates meeting the criteria in Table 1 were shortlisted and interviewed by telephone to check eligibility, namely, older than 18 years; fluency in English; the capacity to contribute to jury discussions; not a health care professional; at least a year as a resident of Greater Manchester; and no special knowledge, interest, or conflict of interest in the jury mission. Jurors were thus chosen to ensure that they had "no special axe to grind" [21].

Reserves attended the jury meeting and stayed until lunchtime on day 1. In each jury, one was needed to replace a juror who did not attend or who had left during the first morning. Both jurors and reserves were paid for their time. One person withdrew from each jury at the end of day 1 for personal reasons. As the reserves had not attended for the afternoon of day 1, and therefore, had not heard the information presented by the witnesses, these jurors were not replaced, leaving 17 people to complete each jury.

Jury Process

The jury mission was planned, designed, and refined over a period of 9 months by a project board comprising five of the authors. The jury mission asked jurors to suppose that a National Health Service (NHS) body wanted to create new records by linking data from the patient records stored by their general practice and by hospitals that have treated them. The new records were for purposes other than direct patient care, including research and service improvement. The jurors were then asked whether this should be allowed and, if so, who should be allowed access to the data. The mission was developed iteratively by the project board to reflect the question on the extent to which patients should control access to patient records (Textbox 1). Both 3-day juries followed the same program (Textbox 3). The activities were designed primarily by the Jefferson Center in line with their citizens' jury method [13] and were managed by two facilitators who were independent of the project board and jury sponsors (KB and AH).

Five expert witnesses were chosen to provide relevant information to and answer any questions from the members of the jury (Table 2). Two witnesses were selected to provide impartial information on day 1, including about the use of deidentified data using the Information Commissioner's Office's (ICO's) anonymization code of practice [22]. The code emphasizes that understanding anonymization means understanding what personal data is, that it can be impossible to assess reidentification risk with absolute certainty, and that different forms of access to anonymized data can pose different reidentification risks (eg, publication is more risky than limited access). Three advocates, known as partial witnesses, were chosen to provide arguments for and against the greater use of patient records on day 2. The purpose of the expert witness presentations was "to inform and educate the jurors, a microcosm of the public, to enable them to reach wise and thoughtful conclusions" [13], rather than to produce. Impartial expert witnesses were asked to confine their presentations and answers to questions to matters of fact rather than values. Partial expert witnesses were asked to make the case for a particular viewpoint or viewpoints based on both facts and values, and an ethicist was asked to provide arguments pulling in both directions (Table 2; see Multimedia Appendix 1 for presentations). The difference between the two types of witnesses was explained to the jurors. After each presentation, there was an opportunity for questions.

Textbox 3. The program of activities for both citizens' juries.

Day 1:

- Participants complete the start-of-Jury questionnaire and consent form
- Introduction to the event
- Group work simulation exercise (about allocation of ambulance services)
- Presentation and questions with expert witness on patient records (Ralph Sullivan), and group work to identify key learning points
- Presentation and questions with expert witness on the law (Dawn Monaghan), and group work to identify key learning points

Day 2:

- Presentation and questions with expert witness arguing for greater use of patient records in the public interest (John Ainsworth), and group work to identify key learning points
- Presentation and questions with expert witness arguing for protection and patient control of patient records (Sam Smith), and group work to identify key learning points
- Presentation and questions with expert witness identifying ethical considerations (Søren Holm), and group work to identify key learning points
- Group work to identify, discuss, and rank reasons for and against the different components of question 1 of the jury mission
- Juror voting on question 1

Day 3:

- Group work with prepared information to develop the case for and against different parties gaining access to records, as set out in question 2 of the jury mission
- Group work to identify, discus, and rank reasons for and against the different parties identified in question 2 of the jury mission
- Juror voting on question 2
- Participants complete the end-of-jury questionnaire

Table 2. Perspectives taken and information provided by impartial and partial witnesses who presented to both juries.

Witnesses	Perspective taken	Information provided
Impartial witnesses		
Dr Ralph Sullivan, general practitioner and medical informatician	To explain what is in a patient record, and how patient records are used in the NHS.	General Medical Council requirements for record keeping, content of multiple patient records, and how they are used in practice both for direct care and secondary uses
Dawn Monaghan, group manager for public services at the Information Commissioner's Office	To tell jurors a little about the law that protects access to patient records.	Outline of relevant privacy law, (common law duty of confidence and Data Protection Act 1998), how data are protected, and limitations to access to data
Partial witnesses		
Dr John Ainsworth, senior research fellow at the University of Manchester	To argue that it's important that patient records are used for research and other purposes that bring benefits to the public.	How data are used to create medical evidence as to the effectiveness and safety of treatment in the public interest
Sam Smith, medConfidential coordinator	Oo make the case for stronger control over access to patient records and better information and choices for patients about the use of patient records.	Risks of reidentification, differences between opt out and opt in, uses of data for decommissioning services, and misuse by commercial companies. Argued the case for greater control of patient records
Professor Søren Holm, professor of bioethics at the University of Manchester	Ethical arguments for patients controlling access to patient records, and ethical arguments for wider use of patient records for the benefit of the public.	Potential benefits of sharing data, problems with sharing data, and difficulties with specific informed consent models. How these conflicting interests can be reconciled. Identified ethical considerations both for patients sharing and for patients controlling patient records for uses other than direct patient care

Jury deliberations occurred in small groups after each presentation and before the preparation of each section of the final report. The small groups recorded and reported the results of their deliberations back to the entire jury. During this time, there were opportunities to seek clarification on points of fact from the experts. In addition, if points had been misunderstood by individual jury members, other jurors corrected them. Over half of the total jury time was devoted to jury deliberations in small groups or together as a large group.

To monitor and minimize bias, an independent oversight panel was appointed. The panel members were chosen from national organizations for their subject knowledge and lack of conflict of interest: the chair of the Confidentiality Advisory Group [23], the assistant director of the Nuffield Council on Bioethics [24], and a senior policy officer from the ICO [25] with responsibility for health data. The panel reviewed the citizens' jury design, the choice of expert witnesses, and much of the detailed jury documentation, including the jury questionnaires and the slides from the presentations by the impartial expert witnesses, resulting in some changes to these materials.

Additional design controls used to monitor and minimize bias included that the project board was only able to influence the jury mission and was independent from the jury process and outcomes. A day-long pilot workshop was conducted with seven members of the public to test aspects of the jury design, including presentations by two of the expert witnesses, some of the planned jury activities, and the pre-and postquestionnaires. This highlighted a number of issues, leading to design changes. During the two juries, jury members were asked to complete a questionnaire at the end of each day as to whether the jury facilitators or anyone else had tried to influence them toward particular conclusions. Paired juries were conducted to reduce bias and validate outcome [18]. Finally, the detailed jury design and results documentation were published online [26].

Jury Questionnaires and Reports

Jurors were asked to complete a questionnaire at the start of the jury to identify their prior views and again after all the jury deliberations were complete (Tables 3 and 4). Data were entered into Excel (Microsoft) and collated using simple counts [13].

During the second half of the jury proceedings, the lead facilitator constructed the juries' report with each jury. The two juries voted on individual aspects of both of the jury mission questions (Textbox 1 and Table 3). Jurors also suggested reasons for and against the jury mission options, and the most important reasons given were chosen by juror voting. Each juror had three votes that could be allocated to two or three of the reasons (no reason could get all three votes). This voting method is now the standard approach of the Jefferson Center, although not described in the manual published in 2004 [13]. It allows jurors to choose more than a single option, which is often desired when faced with a large number of possible selections. These votes and ranked reasons formed the basis of the jury reports. On the afternoons of day 2 and day 3, the facilitator led the jurors through the jury report displayed on a screen, editing in real time in discussion with the jurors to gain their acceptance that it fairly represented their views.

All jurors and reserves consented in writing for the main group deliberations to be video recorded; small group deliberations were not recorded. Eleven jurors from jury 1 also consented to be interviewed briefly on video about their views on the jury mission, to be used in a video about the jury findings [26]. Each interview lasted approximately 3 min. Sections of the videos pertaining to decision making, and the jury report were watched repeatedly by the lead author (MPT). These discussions were

compared with the final versions of the two jury reports that had been prepared contemporaneously with the discussions. Relevant portions of the videoed discussions were transcribed, and verbatim quotes were selected for inclusion to highlight the discussion content. Additional explanations are provided in the quotes inside square brackets, where needed for clarity.

Results

Jury Process

The majority of jurors reported in their postjury questionnaires that there was no evidence of bias in the conduct of the juries. However, bias was reported by a few jurors, particularly regarding what they perceived as the impartiality of information from expert witnesses. Differences in quality of presentations by the witnesses were interpreted by one juror as a deliberate attempt to manipulate proceedings. In addition, one jury member explained in their questionnaire:

> *The roles of the expert witnesses made them naturally inclined to imply certain things, although nothing was explicitly said to persuade us.*

The majority of jurors reported participation in the jury process to be very interesting (12 in jury 1 and 17 in jury 2) or mostly interesting (4 members of jury 1). Throughout, the jury members were fully engaged in the process of deliberation, as was evident from the videos of the proceedings, and the quality of the report that the jurors and facilitators produced.

Jury Questionnaires

Jurors completed the pre- and postjury questionnaires individually (Tables 3 and 4). In jury 1, although 8 jurors did not change their views, 9 jurors did, with 5 of them making shifts in a way that favored public benefits over privacy. In jury 2, although 7 jurors did not change their views at all, 10 did, although the shifts were not as marked as for jury 1. Although 6 jurors moved toward favoring public benefit, 2 moved more toward favoring increased privacy. Figure 1 highlights the changes in opinions of jurors.

At the end of the juries, 33 out of 34 jurors voted independently in support of the secondary use of data, with 24 wanting individuals to be able to opt out and 6 favoring opt-in arrangements (Table 3). The remaining 3 wanted data users only to publish their intentions, with no opportunity for either opting in or out. The *other* suggestions that were given by the jurors were additional requirements as to what data users should be required to do. These included giving opt-out options to children at 16 years and requiring an additional strong regulatory body. One juror wrote on their questionnaire, “I feel if it was an opt-out, people or organisations would just brush over it. Whereas if they want the numbers, it will have to be thought about and [the] public educated.” The reasons for opinion changes were not explicitly ascertained from the jurors, but some volunteered information. One juror expressed new concerns and suspicions as to the rationale for conducting data linkage and the role of the citizens’ jury in giving legitimacy to that process.

Many jurors changed their opinion about who should get access to these records, with more people supporting information sharing to a wider group of people by the end of day 3 (Table 4). The aggregate numbers in the Table belie the fact that individual jurors changed their minds in opposite directions. Four jurors in jury 1 changed their minds about NHS researchers accessing data, with 2 agreeing prejury but not postjury, and 2 agreed to allow access postjury but not prejury. Similarly in jury 2, 6 jurors changed their view regarding both university researchers and local authorities, with 3 moving from denying to allowing access and the other 3 moving in the opposite direction.

Table 3. Results from pre- and postjury questionnaires for jury mission question 1 completed individually by jurors, including changes in opinions. “Change” indicates previous answer to new answer. NHS: National Health Service.

Question and answer options	Jury 1 (n)				Jury 2 (n)			
	Prejury	Change	Postjury	Change	Prejury	Change	Postjury	Change
Should the NHS body be allowed to create these records about you and other patients?								
a. Yes, but they should publish information about what they plan to do	2	a→b (2)	2	b→a (2)	0		2	b→a (2)
b. Yes, but they should publish information about what they plan to do and patients should be able to opt out	8	b→a (2); b→b (6)	13	a→b (2); b→b (6); c→b (5)	12	b→a (2); b→b (7); b→c (2); b→e (1)	10	b→b (7); c→b (3)
c. Yes, but they should publish information about what they plan to do and only create records for patients who opt in	6	c→b (5); c→c (1)	1	c→c (1)	4	c→b (3); c→e (1)	3	b→c (2); e→c (1)
d. No	0		0		0		0	
e. Other	1	e→e (1)	1	e→e (1)	1	e→c (1)	2	b→e (1); c→e (1)

Table 4. Results from pre- and post-jury questionnaires for jury mission question 2 completed individually by jurors, including changes in opinions. "Change" indicates previous answer to new answer. Y=organization should be granted access; N=organization should not be granted access. NHS: National Health Service.

Question and answer options	Jury 1 (n)				Jury 2 (n)			
	Prejury	Change	Postjury	Change	Prejury	Change	Postjury	Change
Which organizations should be granted access to these records? (Choose all that apply)								
NHS clinicians and administrators who decide which health services should (and should not) be funded								
Yes	10	Y→Y (10)	15	Y→Y (10); N→Y (5)	7	Y→Y (7)	17	Y→Y (7); N→Y (10)
No	7	N→Y (5); N→N (2)	2	N→N (2)	10	N→Y (10)	0	
NHS clinicians and administrators doing approved research into whether doctors are prescribing medicines appropriately								
Yes	15	Y→Y (13); Y→N (2)	15	Y→Y (13); N→Y (2)	14	Y→Y (14)	17	Y→Y (14); N→Y (3)
No	2	N→Y (2)	2	Y→N (2)	3	N→Y (3)	0	
University staff doing approved research into whether doctors are prescribing medicines appropriately								
Yes	9	Y→Y (8); Y→N (1)	15	Y→Y (8); N→Y (7)	14	Y→Y (11); Y→N (3)	14	Y→Y (11); N→Y (3)
No	8	N→Y (7); N→N (1)	2	Y→N (1); N→N (1)	3	N→Y (3)	3	Y→N (3)
Staff employed by local authorities planning the future need for residential care homes								
Yes	4	Y→Y (3); Y→N (1)	10	Y→Y (3); N→Y (7)	6	Y→Y (3); Y→N (3)	6	Y→Y (3); N→Y (3)
No	13	N→Y (7); N→N (6)	7	N→N (6); Y→N (1)	11	N→Y (3); N→N (8)	11	N→N (8); Y→N (3)
Staff employed by a private company being paid by a hospital NHS trust to compare the number of people dying after surgery with other hospitals								
Yes	5	Y→Y (5)	10	Y→Y (5); N→Y (5)	1	Y→Y (1)	6	Y→Y (1); N→Y (5)
No	12	N→Y (5); N→N (7)	7	N→N (7)	16	N→Y (5); N→N (11)	11	N→N (11)
Staff employed by an insurance company aiming to set health insurance premiums accurately								
Yes	2	Y→Y (1); Y→N (1)	3	Y→Y (1); N→Y (2)	0		1	N→Y (1)
No	15	N→Y (2); N→N (13)	14	N→N (13); Y→N (1)	17	N→Y (1); N→N (16)	16	N→N (16)
Staff employed by a pharmaceutical company investigating whether they should begin research into a new drug for a genetic disease for which there is currently no treatment								
Yes	7	Y→Y (7)	12	Y→Y (7); N→Y (5)	5	Y→Y (4); Y→N (1)	10	Y→Y (4); N→Y (6)
No	10	N→Y (5); N→N (5)	5	N→N (5)	12	N→Y (6); N→N (6)	7	N→N (6); Y→N (1)

Figure 1. Numbers of jurors who changed their answers to question 1 of the jury mission.

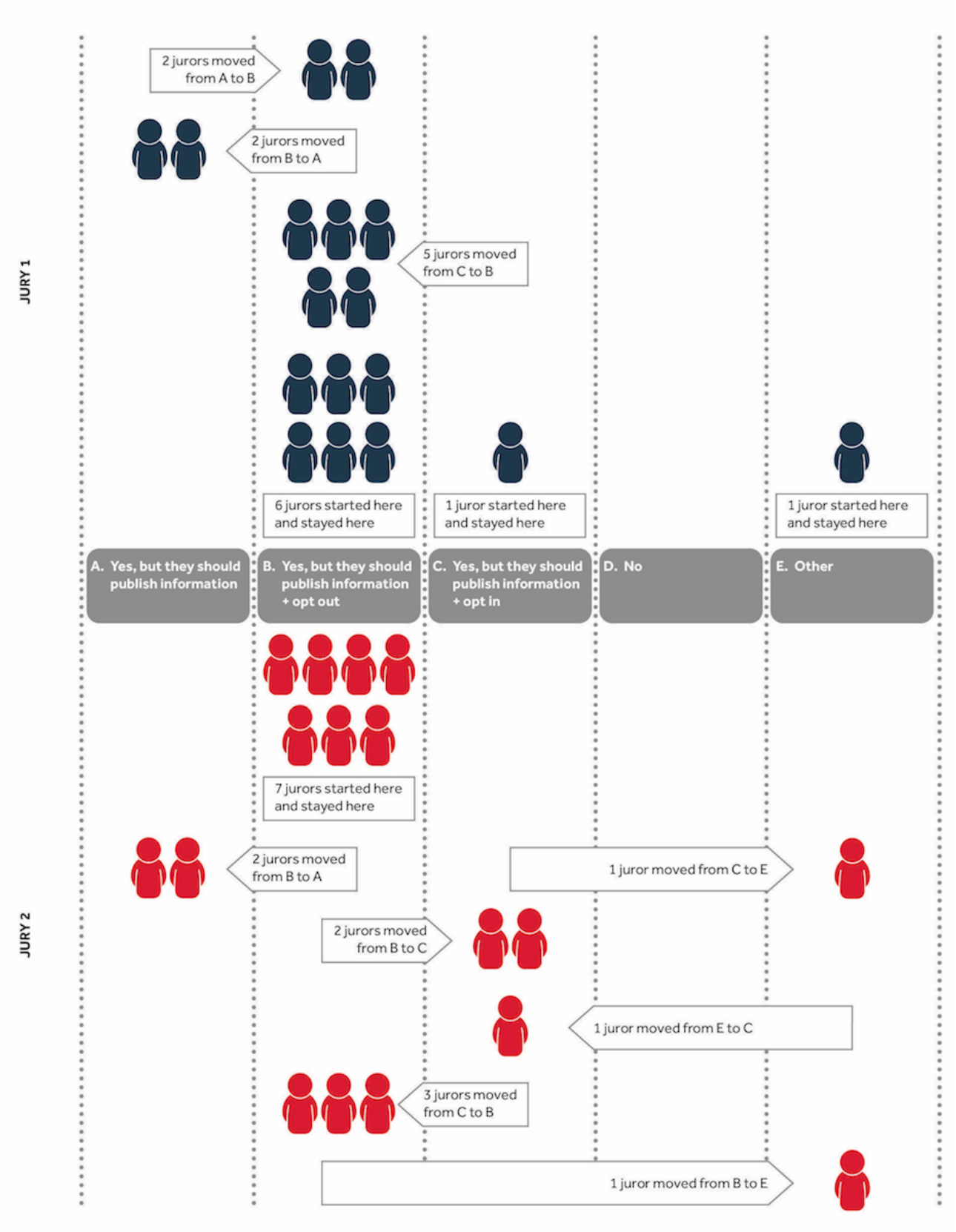


Table 5. Votes on subquestions of the jury mission completed during the writing of the jury report. NHS: National Health Service.

Questions and answer options	Jury 1	Jury 2
Vote 1A. Should the NHS body be allowed to create these records about you and other patients?		
Yes	17	13
No	0	4
Vote 1B. If such records were created, should they only publish information about what they plan to do or allow a patient option (type unspecified)		
Publish only	8	5
Patient option	9	12
Vote 1C. Should individuals have the option to opt in or opt out?		
Opt in	1	5
Opt out	16	12

Jury Reports

In addition to the questionnaires above, jurors voted separately on three individual aspects of question 1, by private poll, during the writing of the report with the facilitator (Table 5). Voting on each subquestion was undertaken over the afternoon of day 2 (vote 1A) and throughout day 3 (votes 1B and 1C). As a consequence, jurors could and did change their opinions during later votes without being able to change the vote they had given for earlier questions. Thus, in jury 1, for example, only 9 jurors voted for patient input in vote 1B, but all were required to choose between two different types of patient input in vote 1C. The reasons for changing their minds could be either based on individual rights or pragmatism. Two jury members from jury 1 stated:

> *I changed my mind following this discussion, yeah, well I would have said that they should just announce it [publish] because I thought if they are going to do it [use data], do it so it is accurate and everybody is included in it. But now I am thinking, well human rights, should people have a decision, you know, whether to be included or not, yeah.*

> *I was thinking publish but, in reality, if we give people an option, I don't think they will opt out as much, but if you don't give them an option most people will go off [get angry].*

Use of Data

In vote 1A, all jurors in jury 1, and all but 4 jurors in jury 2, voted that the NHS should be allowed to create linked records (Table 5). The reason most commonly voted for by both juries was that more detailed and complete data would produce more accurate evidence, which can lead to more effective, more cost-effective health care through the NHS (see Multimedia Appendix 2). Other reasons included that personalized medicine and treatments could only be discovered and used effectively through use of more complete data and records and that data use would help identify ineffective drugs and treatments sooner so that they can be removed from use and increase patient safety.

The main argument mentioned by most jury members against creating and sharing linked patient records (regardless of whether they voted in favor of still doing it) concerned transparency of use. They felt that, without a clear understanding of who would be regulating the data and making decisions about access, it was difficult to support the creation of new records. In addition, there were concerns that this would benefit researchers or companies rather than individual patients, as explained by this juror from jury 2:

> *What proof is there that the general public would be any better off in terms of that research with all the data being in one place, then? Whereas at the moment there is all this research going on anyway, but it is more difficult for the researchers and the private companies to get the information, because they have to go all over for it. So is there any evidence to prove that it would be better than what it is at the moment?*

There were concerns that despite safeguards, data and records may not be secure and may be accessed by individuals or organizations without proper permission or legal authority, or for reasons other than where originally authorized. In addition, there were concerns that data may be used by private companies for commercial gain rather than for the benefit of patients and the public, or sold on to other companies.

Patient Choice

The issue of whether there should be patient choice (either opt out or opt in), or whether the NHS body should merely follow the minimum legal requirements of publishing that data use had taken place, divided both juries. In jury 1, 8 people voted to publish only, whereas in jury 2, only 5 people voted for this. The reasons given were that this would ensure more accurate, complete data when all records are included, which would be of greater benefit to the population and that it would save time and money through a much more streamlined, efficient process. Reasons given why patients should have input included that there was an expectation that people should be able to have autonomy and freedom of choice by having control over their own data and records and choosing whether or not their record is included and that it would allow individuals to maintain their confidence in doctors and other health care settings where trust is critical. In addition, it was suggested that the process of obtaining patient input would allow greater transparency in how records are used and shared.

Jury 1 had more members suggesting an opt-out model in comparison with jury 2. The most frequently selected reason they gave for suggesting this model was that more people would be included in the data, and this would lead to more accurate results and more representative samples of the population, and this in the end would lead to more rigorous research and better treatments. Other reasons included that this would be more effective in terms of time and money, as it was an easier and more convenient option for individuals. The alternative would take an enormous effort and may still not properly provide the opportunity to every individual to make an informed decision. This option would allow those who may simply be undecided (but not opposed) to still contribute to research and improvements in health care. The reasons that jurors suggested for an opt-in option included that this option would require the organization to conduct an information campaign to educate the public and would mean that individuals whose data were used in analysis could make an informed decision to be included.

Data Access

When considering who should get access to data, the two juries had very similar rationales, which were written in the report that they produced at the end of the third day. Both thought that public benefit was a key justification for access. Jury 1 was more strongly supportive of sharing patient records for public benefit, whereas jury 2 was more cautious and sought to give patients more control. In particular, they concluded that organizations and individuals who *should* be granted access to these records tend to demonstrate similar characteristics. Typically, these organizations clearly demonstrated that the primary goal for using the data was for public benefit (such as improved medical care and treatments, improved public health, or management of public funds) and made a clear and compelling case for why they need these patient records. They provided clear justification for how and why the data would be used, why it was relevant to their efforts, with whom it will be shared, and only access records they needed to perform their data analysis and could not get adequate data from other sources. The organizations showed a clear, relevant connection between the issues they are addressing and the information contained in these records, had a track record of protecting data and records, and could be trusted to maintain control of data without sharing and have controls in place to properly secure the data and safeguard against internal misuse. Finally, these organizations needed access to the data to conduct urgent and/or timely analysis.

The reasons the juries gave for why organizations should not have access to the data included several that were the opposite of the reasons for access, such as organizations that did not clearly indicate that the primary use of the data is for public benefit, who may use the data solely for private gain or commercial profit, or who did not have a trusted track record for protecting data. In addition, they might use the data to exploit or manipulate individuals or populations or might manipulate the data to support their own agenda.

Discussion

Principal Findings

The findings of the citizens' jury work highlight that, when informed of both the risks and opportunities associated with health data sharing, members of the public believe an individual's right to privacy should not prevent research that can benefit patients overall. The juries also concluded that patients should be notified of any such scheme and have the right to opt out if they so choose. Many, but not all, jurors became less skeptical about health data sharing, as they became better informed of its benefits and risks.

Attitudes to Data Use

The findings from this study support the contention that some members of the public believe that NHS records are a public resource, paid for by public money, and therefore, should be used for research for the public benefit [27,28]. Few jurors objected to the use of health data *per se*, but many wanted, as a minimum, to be told that such uses were happening and to be given an option to opt out. It reinforces the fact that the social license, or the societal expectations as to how deidentified data should be used, is not necessarily the same as what is permissible by law [4].

There were both individual and aggregate changes in attitude, which has been found in some [8] but not all [29,30] previous studies in this area. The jury members had the opportunity to learn about and deliberate on the general use of linked health data over the course of 3 days, which may well have contributed to how they changed their opinions. For most of the jury members, this change related to becoming more accepting of less patient control over the use of data or more pragmatic about the need for slightly more patient control, depending on their initial views. However, two jury members changed their mind quite strikingly and became much more insistent upon greater patient control.

The reasons for opinion changes were not explicitly ascertained in this study. Other studies have found that the provision of general information [31] or information about the impact of selection bias [8] may be important in changing opinions toward greater acceptance of use of deidentified data without explicit consent. Other studies have shown that, during focus groups, people change their mind to become more accepting about such data use, rather than the reverse [11,32]. However, this study also found that, for a minority of people, their opinion changed toward being more skeptical about data use. From the perspective of public engagement about data use and linkage, this suggests that some individuals may well receive the same information but reach different conclusions to their peers, perhaps by applying different values. It also calls into question the assumption that public distrust will necessarily be addressed, such as the deficit model of public understanding of science, by simply providing greater dissemination of information [33,34].

Many of the jurors changed their minds from preferring either no public input or an opt-in model to preferring an opt-out model. Hill and colleagues found a lack of consensus in the

international literature on a model of consent that was preferred by the public [8]. Taylor and Taylor found, in their small-scale study, that although some people may well *prefer* opt-in models, pragmatically, they would be willing to *accept* opt-out models [29]. Our study did not test other, more nuanced, models of consent, such as dynamic consent [35], which enable people to amend their choices as often as they wish, when they change their minds as to what they are willing to permit to happen with data about them.

The jurors wanted data to only be provided to organizations that could demonstrate that the primary goal for using the data was for public benefit (either for new treatments or to improve existing services). In addition, such organizations would have to be trustworthy because of their previous track record and existing controls and safeguards against misuse. This is similar to what has been found elsewhere, both nationally [11,36] and internationally [9,37]. Although commercially funded research has been considered unacceptable in some studies [8], some jurors became more willing to accept such uses by the end of the jury proceedings. This may reflect that *why* commercial research was being conducted (eg, for public benefit) mattered more than who was conducting it [11].

Use of Citizens' Juries

As has been found in other studies [38], this work shows that citizens are capable of critically evaluating expert opinion presented to them, identifying and seeking out any additional knowledge they need by asking questions from the witnesses, and then using deliberation to reach an agreed opinion. Such deliberation went on throughout the jury process, as was seen when jury members described changing opinions following discussions with their peers and the differences seen in the prejury and postjury questionnaires.

These citizens' juries were conducted as close as possible to the ideal suggested in a recent systematic review [18]. Our links with the Jefferson Center and having a member of their staff act as lead jury facilitator (KB) ensured that the citizens' juries demonstrated the three important characteristics of deliberative democracy: inclusivity, deliberation, and active citizenship [18]. The selection criteria ensured that people from a broad range of backgrounds were recruited, including those with opposing opinions on privacy and whose voices might not otherwise be heard [39]. The recruitment method was different from that found in the Jefferson Center manual, which advocated cold calling random telephone numbers [13]. Even in 2004, the authors cautioned that 180 calls would be needed to recruit each juror. Online recruitment and completion of a screening questionnaire was a 21st century update (approved by the Jefferson Center) that proved much more efficient, as it required little staff time to identify potential jurors.

By paying jurors appropriately for their time, it was possible to ensure that participants were representative of the population and not limited to the subgroup of people with sufficient resources to enable them to engage in a lengthy volunteer activity. Citizens' juries use panels that are selected to be representative of the population [13]. The Jefferson Center manual highlights that criteria should be both demographic and attitudinal, with targets based on existing data. Hence, we used a national census for criteria based on demographic data [19] and a recent survey of public opinion on balancing privacy against information sharing for privacy base criteria [20]. We acknowledge that this produced a jury with more people in favor of data sharing, but this reflects the views of the overall British population. In a small sample of 18 people for each jury, it was important that we did not by chance recruit a disproportionate number of people who were very supportive of information sharing, or a disproportionate number of people who were very privacy conscious and cautious about information sharing.

The witnesses presented diverse viewpoints, and adequate time was allowed for jurors to question and challenge the witnesses. More than half the available time was given to jury deliberation, which was conducted in small groups with varying participants, to ensure that jurors interacted with all others. The evidence from the videos shows the care with which the jurors approached their role. The jurors were told from the beginning that the findings from the juries would be fed back to policy makers to ensure active citizenship. Twelve jurors (6 jurors from each jury) were invited to the postjury workshop where they successfully engaged with the invited stakeholders, including national policy makers.

The findings from citizens' juries are qualitative in nature, and therefore, the findings are not intended to be generalizable in a statistical sense. As described by Lincoln and Guba [40], qualitative research aims for transferability by showing how the findings may be applicable to other contexts. To achieve this, the jury process and materials have been published on the Web [26], alongside this paper, to increase transparency and allow other readers to consider whether the findings are applicable in broader contexts.

Bias, both conscious and unconscious, is an important criticism of citizens' juries [41]. Despite the efforts of the researchers and the independent oversight panel, to ensure that jurors were presented with balanced information to ensure overall fairness, some jurors reported a perception of bias. The witnesses had been chosen so that one set were intentionally impartial and one set were intentionally partial and aiming to present a particular side of the arguments to the jurors relative to the mission. There may have been a lack of clarity about this for several jurors, which could have led to these impressions. This suggests that bias in citizens' juries can be monitored and minimized, but not totally eliminated.

It could be argued that deliberative methods should be used for making complex policy decisions. There is some evidence from the literature that people are happy to have an ethics committee make decisions about whether to approve the use of data for individual research studies [42]. In addition, there have been centuries of experience of using a 12-person jury in criminal trials. It has been suggested that citizens' juries *symbolically* represent the community [41,43]. Nonetheless, citizens' juries such as the ones conducted here are not usually given public accountability for their decisions and therefore, may be less acceptable to members of the general public.

From a practical perspective, however, deliberative methods such as citizens' juries are not a reasonable choice for all policy decisions because they are so resource-intensive. They could

be a good choice for situations where the topics are both important and potentially intractable or where democratic legitimacy is needed for decisions [44]. In addition, there is no expectation that the methodology should (or indeed could) be scaled up to provide a large-scale public engagement activity itself. Alternative activities that are designed to reach large numbers of citizens, however, can be informed by the knowledge gained from having previously conducted citizens' juries.

Conclusions

Our citizens' jury method was successful in enabling members of the public to deliberate and make decisions about a complex policy problem. Many jurors became less skeptical about health data sharing, as they became better informed of its benefits and risks. Most jurors wanted public input in the form of information provision and the right to opt out. This was one of only a few studies to show that during a deliberative process, a small minority of people become more skeptical about data use, rather than less. This suggests that public engagement about the data use cannot assume that merely providing more public information will equal more public trust.

The deliberative method used in this study may help uncover often-overlooked opportunities for policy makers to engage meaningfully and substantively with the public about technical, and potentially divisive, public policy issues—especially those that have been recently controversial. This research demonstrates that citizens' juries can be an effective model for engaging the public on policy issues that balance competing issues such as potential risks to individuals, the pursuit of commercial profit, the search for answers to research questions by academic institutions, and the possibility of direct public benefit for society as a whole. Further research is needed as to whether citizens' juries would be acceptable to the public as a way to have an informed set of peers make decisions on their behalf.

Acknowledgments

The work described in this paper was part-funded by the National Institute for Health Research (NIHR), Greater Manchester Primary Care Patient Safety Translational Research Centre (Greater Manchester PSTRC), and also from The Farr Institute @ HeRC. The Farr Institute @ HeRC is supported by a 10-funder consortium: Arthritis Research UK, the British Heart Foundation, Cancer Research UK, the Economic and Social Research Council, the Engineering and Physical Sciences Research Council, the Medical Research Council, the NIHR, the National Institute for Social Care and Health Research (Welsh Assembly Government), the Chief Scientist Office (Scottish Government Health Directorates), and the Wellcome Trust, (MRC Grant No: MR/K006665/1). The views expressed are those of the authors and not necessarily those of the NHS, the Department of Health, the NIHR, or any of the funders of The Farr Institute.

Conflicts of Interest

KB is Executive Director of the Jefferson Center, which invented the citizens' jury process. MO is Director of Citizens' Juries Community Interest Company, a social enterprise dedicated to designing and running citizens' juries, and was commissioned to deliver these juries.

Multimedia Appendix 1

Presentations from expert witnesses.

[PDF File (Adobe PDF File), 6MB - jmir_v20i3e112_app1.pdf]

Multimedia Appendix 2

Main reasons given by juries for their votes for part 1 of jury mission from the jury reports.

[PDF File (Adobe PDF File), 34KB - jmir_v20i3e112_app2.pdf]

References

1. John A, Marchant AL, McGregor JI, Tan JO, Hutchings HA, Kovess V, et al. Recent trends in the incidence of anxiety and prescription of anxiolytics and hypnotics in children and young people: An e-cohort study. J Affect Disord 2015 Sep 01;183:134-141 [FREE Full text] [doi: 10.1016/j.jad.2015.05.002] [Medline: 26005207]
2. Bergman BP, Mackay DF, Smith DJ, Pell JP. Long-term mental health outcomes of military service: national linkage study of 57,000 veterans and 173,000 matched nonveterans. J Clin Psychiatry 2016 Jun;77(6):793-798. [doi: 10.4088/JCP.15m09837] [Medline: 27135139]
3. Mackay DF, Haw S, Newby DE, Langhorne P, Lloyd SM, McConnachie A, et al. Impact of Scotland's comprehensive, smoke-free legislation on stroke. PLoS One 2013 May;8(5):e62597 [FREE Full text] [doi: 10.1371/journal.pone.0062597] [Medline: 23667497]
4. Carter P, Laurie GT, Dixon-Woods M. The social licence for research: why care.data ran into trouble. J Med Ethics 2015 May;41(5):404-409 [FREE Full text] [doi: 10.1136/medethics-2014-102374] [Medline: 25617016]

5. Garrety K, McLoughlin I, Wilson R, Zelle G, Martin M. National electronic health records and the digital disruption of moral orders. Soc Sci Med 2014 Jan;101:70-77. [doi: 10.1016/j.socscimed.2013.11.029] [Medline: 24560226]
6. Kho ME, Duffett M, Willison DJ, Cook DJ, Brouwers MC. Written informed consent and selection bias in observational studies using medical records: systematic review. Br Med J 2009 Mar 12;338:b866. [doi: 10.1136/bmj.b866]
7. Davies GM, Robinson M, Neville J, Burnside G. Investigation of bias related to non-return of consent for a dental epidemiological survey of caries among five-year-olds. Community Dent Health 2014 Mar;31(1):21-26. [Medline: 24741889]
8. Hill EM, Turner EL, Martin RM, Donovan JL. "Let's get the best quality research we can": public awareness and acceptance of consent to use existing data in health research: a systematic review and qualitative study. BMC Med Res Methodol 2013;13:72 [FREE Full text] [doi: 10.1186/1471-2288-13-72] [Medline: 23734773]
9. Damschroder LJ, Pritts JL, Neblo MA, Kalarickal RJ, Creswell JW, Hayward RA. Patients, privacy and trust: patients' willingness to allow researchers to access their medical records. Soc Sci Med 2007 Jan;64(1):223-235. [doi: 10.1016/j.socscimed.2006.08.045] [Medline: 17045717]
10. Aitken M, de St. Jorre J, Pagliari C, Jepson R, Cunningham-Burley S. Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies. BMC Med Ethics 2016 Dec 10;17(1):73 [FREE Full text] [doi: 10.1186/s12910-016-0153-x] [Medline: 27832780]
11. Ipsos MORI. Wellcome Trust. 2016. The One-Way Mirror: Public attitudes to commercial access to health data Date URL: https://wellcome.ac.uk/sites/default/files/public-attitudes-to-commercial-access-to-health-data-wellcome-mar16.pdf [accessed 2018-03-08] [WebCite Cache ID 6xlTDsgQW]
12. Aitken M, Cunningham-Burley S, Pagliari C. Moving from trust to trustworthiness: experiences of public engagement in the Scottish Health Informatics Programme. Sci Public Policy 2016 Oct;43(5):713-723 [FREE Full text] [doi: 10.1093/scipol/scv075] [Medline: 28066123]
13. Veasey K, Nethercut D. Citizens Jury Handbook. St Paul, Minnesota: Jefferson Center; 2004. URL: https://jefferson-center.org/wp-content/uploads/2018/03/Jefferson-Center-CJ-Handbook-2004.pdf [WebCite Cache ID 6xlen3znz]
14. Iredale R, Longley M, Thomas C, Shaw A. What choices should we be able to make about designer babies? A citizens' jury of young people in South Wales. Health Expect 2006 Sep;9(3):207-217 [FREE Full text] [doi: 10.1111/j.1369-7625.2006.00387.x] [Medline: 16911135]
15. City of Melbourne. 10 year financial plan URL: https://participate.melbourne.vic.gov.au/10yearplan [WebCite Cache ID 6uQJu188x]
16. Farr Institute of Health Informatics Research. URL: http://www.farrinstitute.org/ [WebCite Cache ID 6xlU5ekIP]
17. The University of Manchester. Greater Manchester Primary Care Patient Safety Translational Research Centre URL: http://www.patientsafety.manchester.ac.uk/ [accessed 2018-03-08] [WebCite Cache ID 6xlUCEp9H]
18. Street J, Duszynski K, Krawczyk S, Braunack-Mayer A. The use of citizens' juries in health policy decision-making: a systematic review. Soc Sci Med 2014 May;109:1-9 [FREE Full text] [doi: 10.1016/j.socscimed.2014.03.005] [Medline: 24657639]
19. Office for National Statistics, National Records of Scotland, Northern Ireland Statistics Research Agency. InFuse. 2011 Census aggregate data URL: http://infuse.ukdataservice.ac.uk/ [accessed 2018-03-08] [WebCite Cache ID 6xlV0kfYI]
20. Royal Statistical Society. 2014. Royal Statistical Society research on trust in data and attitudes toward data use/data sharing URL: https://www.statslife.org.uk/images/pdf/rss-data-trust-data-sharing-attitudes-research-note.pdf [WebCite Cache ID 6uQKsMF3z]
21. Dunkerley D, Glasner P. Empowering the public? Citizens' juries and the new genetic technologies. Crit Public Health 1998 Sep;8(3):181-192. [doi: 10.1080/09581599808402906]
22. Information Commissioners Office. Anonymisation: managing data protection risk code of practice URL: https://ico.org.uk/for-organisations/guide-to-data-protection/anonymisation/ [WebCite Cache ID 6uVKws7MH]
23. Health Research Authority. Section 251 and the Confidentiality Advisory Group (CAG) URL: http://www.hra.nhs.uk/about-the-hra/our-committees/section-251/ [accessed 2017-10-26] [WebCite Cache ID 6uVInjrUU]
24. Nuffield Council for Bioethics. URL: http://nuffieldbioethics.org/about [WebCite Cache ID 6xlfLXHdE]
25. Information Commissioner's Office. URL: https://ico.org.uk/ [WebCite Cache ID 6uVJFv3dj]
26. Health e-Research Centre. 2016. Citizens' juries: health data on trial URL: https://www.herc.ac.uk/get-involved/citizens-jury/ [WebCite Cache ID 6uQKxSn91]
27. Cassell J, Young A. Why we should not seek individual informed consent for participation in health services research. J Med Ethics 2002;28:313-317. [doi: 10.1136/jme.28.5.313]
28. Miller FG. Research on medical records without informed consent. J Law Med Ethics 2008;36(3):560-566. [doi: 10.1111/j.1748-720X.2008.304.x] [Medline: 18840249]
29. Taylor MJ, Taylor N. Health research access to personal confidential data in England and Wales: assessing any gap in public attitude between preferable and acceptable models of consent. Life Sci Soc Policy 2014 Dec;10:15 [FREE Full text] [doi: 10.1186/s40504-014-0015-6] [Medline: 26085451]
30. Willison DJ, Swinton M, Schwartz L, Abelson J, Charles C, Northrup D, et al. Alternatives to project-specific consent for access to personal information for health research: insights from a public dialogue. BMC Med Ethics 2008 Nov 19;9:18 [FREE Full text] [doi: 10.1186/1472-6939-9-18] [Medline: 19019239]

31. Medical Research Council. 2007. The Use of Personal Health Information in Medical Research URL: https://www.mrc.ac.uk/documents/pdf/the-use-of-personal-health-information-in-medical-research-june-2007/ [accessed 2018-03-08] [WebCite Cache ID 6xlfihgQu]
32. Nair K, Willison D, Holbrook A, Keshavjee K. Patients' consent preferences regarding the use of their health information for research purposes: a qualitative study. J Health Serv Res Policy 2004 Jan;9(1):22-27. [doi: 10.1258/135581904322716076] [Medline: 15006236]
33. Pouliot C, Godbout J. Thinking outside the 'knowledge deficit' box. EMBO Rep 2014 Aug;15(8):833-835 [FREE Full text] [doi: 10.15252/embr.201438590] [Medline: 24993560]
34. Cunningham-Burley S. Public knowledge and public trust. Community Genet 2006 Jun;9(3):204-210. [doi: 10.1159/000092658] [Medline: 16741351]
35. Williams H, Spencer K, Sanders C, Lund D, Whitley EA, Kaye J, et al. Dynamic consent: a possible solution to improve patient confidence and trust in how electronic patient records are used in medical research. JMIR Med Inform 2015;3(1):e3 [FREE Full text] [doi: 10.2196/medinform.3525] [Medline: 25586934]
36. NHS Information Authority, The Consumers' Association. Share with care! People's views on consent and confidentiality of patient information. Birmingham: NHS Information Authority; 2002.
37. Willison DJ, Schwartz L, Abelson J, Charles C, Swinton M, Northrup D, et al. Alternatives to project-specific consent for access to personal information for health research: what is the opinion of the Canadian public? J Am Med Inform Assoc 2007;14(6):706-712 [FREE Full text] [doi: 10.1197/jamia.M2457] [Medline: 17712084]
38. Bennett P, Smith SJ. Genetics, insurance and participation: how a citizens' jury reached its verdict. Soc Sci Med 2007 Jun;64(12):2487-2498. [doi: 10.1016/j.socscimed.2007.02.029] [Medline: 17418469]
39. Menon D, Stafinski T. Engaging the public in priority-setting for health technology assessment: findings from a citizens' jury. Health Expect 2008 Sep;11(3):282-293 [FREE Full text] [doi: 10.1111/j.1369-7625.2008.00501.x] [Medline: 18816324]
40. Lincoln YS, Guba EG. Naturalistic Inquiry. Newbury Park, CA: Sage Publications; 1985.
41. Smith G, Wales C. The theory and practice of citizens' juries. Policy Polit 1999 Jul;27(3):295-308. [doi: 10.1332/030557399782453118]
42. Haddow G, Bruce A, Sathanandam S, Wyatt JC. 'Nothing is really safe': a focus group study on the processes of anonymizing and sharing of health data for research purposes. J Eval Clin Pract 2011 Dec;17(6):1140-1146. [doi: 10.1111/j.1365-2753.2010.01488.x] [Medline: 20629997]
43. Lenaghan J. Involving the public in rationing decisions. The experience of citizens juries. Health Policy 1999 Oct;49(1-2):45-61. [Medline: 10827290]
44. Dryzek JS. Discursive democracy in a reflexive modernity. In: Dryzek JS, editor. Deliberative Democracy and Beyond: Liberals, Critics, Contestations. Oxford: Oxford University Press; 2000:162-175.

Abbreviations

ICO: Information Commissioner’s Office
NHS: National Health Service

Edited by G Eysenbach; submitted 28.03.17; peer-reviewed by D Willison, A Waldo, S Wilson, L van Velsen; comments to author 12.07.17; revised version received 02.11.17; accepted 16.11.17; published 28.03.18.

Please cite as:
Tully MP, Bozentko K, Clement S, Hunn A, Hassan L, Norris R, Oswald M, Peek N
Investigating the Extent to Which Patients Should Control Access to Patient Records for Research: A Deliberative Process Using Citizens’ Juries
J Med Internet Res 2018;20(3):e112
URL: http://www.jmir.org/2018/3/e112/
doi:10.2196/jmir.7763
PMID:29592847

Original Paper

Self-Swabbing for Virological Confirmation of Influenza-Like Illness Among an Internet-Based Cohort in the UK During the 2014-2015 Flu Season: Pilot Study

Clare Wenham[1]; Eleanor R Gray[2]; Candice E Keane[2,3]; Matthew Donati[4]; Daniela Paolotti[5]; Richard Pebody[6]; Ellen Fragaszy[1,7]; Rachel A McKendry[2,3]; W John Edmunds[1]

[1]Department of Infectious Disease Epidemiology, London School of Hygiene and Tropical Medicine, London, United Kingdom
[2]London Centre for Nanotechnology, University College London, London, United Kingdom
[3]Division of Medicine, University College London, London, United Kingdom
[4]Bristol Public Health Laboratory, Public Health England, Bristol, United Kingdom
[5]ISI Foundation, Torino, Italy
[6]Centre of Infectious Disease Surveillance and Control, Public Health England, London, United Kingdom
[7]Centre for Public Health Data Science, Institute of Infectious Disease Informatics, University College London, London, United Kingdom

Corresponding Author:
Clare Wenham
Department of Infectious Disease Epidemiology
London School of Hygiene and Tropical Medicine
Keppel Street
London, WC1E 7HT
United Kingdom
Phone: 44 207 955 ext 6592
Email: c.wenham@lse.ac.uk

Abstract

Background: Routine influenza surveillance, based on laboratory confirmation of viral infection, often fails to estimate the true burden of influenza-like illness (ILI) in the community because those with ILI often manage their own symptoms without visiting a health professional. Internet-based surveillance can complement this traditional surveillance by measuring symptoms and health behavior of a population with minimal time delay. Flusurvey, the UK's largest crowd-sourced platform for surveillance of influenza, collects routine data on more than 6000 voluntary participants and offers real-time estimates of ILI circulation. However, one criticism of this method of surveillance is that it is only able to assess ILI, rather than virologically confirmed influenza.

Objective: We designed a pilot study to see if it was feasible to ask individuals from the Flusurvey platform to perform a self-swabbing task and to assess whether they were able to collect samples with a suitable viral content to detect an influenza virus in the laboratory.

Methods: Virological swabbing kits were sent to pilot study participants, who then monitored their ILI symptoms over the influenza season (2014-2015) through the Flusurvey platform. If they reported ILI, they were asked to undertake self-swabbing and return the swabs to a Public Health England laboratory for multiplex respiratory virus polymerase chain reaction testing.

Results: A total of 700 swab kits were distributed at the start of the study; from these, 66 participants met the definition for ILI and were asked to return samples. In all, 51 samples were received in the laboratory, 18 of which tested positive for a viral cause of ILI (35%).

Conclusions: This demonstrated proof of concept that it is possible to apply self-swabbing for virological laboratory testing to an online cohort study. This pilot does not have significant numbers to validate whether Flusurvey surveillance accurately reflects influenza infection in the community, but highlights that the methodology is feasible. Self-swabbing could be expanded to larger online surveillance activities, such as during the initial stages of a pandemic, to understand community transmission or to better assess interseasonal activity.

(J Med Internet Res 2018;20(3):e71) doi:10.2196/jmir.9084

KEYWORDS

influenza; influenza-like illness; surveillance; online; cohort study; virological confirmation

Introduction

Influenza and influenza-like illness (ILI) cause a considerable burden of illness in the UK [1]. For most people, influenza is usually a self-limited disease for which, on average, between 1.5 and 4.9 working days are lost for each episode [2]. The consequences for high-risk groups (very young, older people, pregnant women, and those with an underlying health condition) can be more serious. Public Health England (PHE) estimate that each winter hundreds of thousands of people see their general practitioner (GP), tens of thousands are hospitalized, and there are on average 8000 deaths because of influenza [3-5]. Moreover, it has been estimated in 2011 that an outbreak of pandemic influenza could reduce the UK gross domestic product by approximately 1.14% to 1.42% (£14.8-£18.5 billion) [6].

Surveillance is an essential function for monitoring seasonal and pandemic influenza, delivering epidemiological, virological, and clinical awareness of the circulating virus and studying interventions, such as vaccination programs [7]. However, routine surveillance through medical settings (GPs and hospitals), based on laboratory confirmation of infection, does not provide a full picture of the true societal burden of influenza at any one time due to the fact that individuals suffering from ILI often do not visit a health care professional, but manage their symptoms on their own [8,9]. Syndromic surveillance of ILI is increasingly used as a method for detecting discernible trends in illness, without laboratory confirmation [10]. For example, PHE collates data from a range of sources to compile its weekly national influenza reports, including NHS 111 (a nonemergency health advice phone line), the Royal College of General Practitioners' (RCGP) Weekly Returns Service (GP-based sentinel surveillance), Medical Offices of Schools Association, community telephone surveys, and online disease surveillance platforms such as Flusurvey [11].

Flusurvey, the UK's largest crowd-sourced platform for surveillance of influenza, collects routine data on more than 6000 voluntary participants [12]. On registration, a baseline epidemiological questionnaire is carried out, asking about individuals' age, gender, location (first part of postcode), household composition, influenza vaccine status, and preexisting health conditions. Although Flusurvey participants are not representative of the UK population, adjustments are made through modeling processes to allow for broader calculations to be made at the population level. Subsequently, participants are emailed each week to complete a symptoms survey. Participants select recent symptoms from a list including respiratory and gastrointestinal concerns, and provide information relating to onset and duration of symptoms and health-seeking behavior, as well as rating how they are feeling overall on a scale of 1 (very unwell) to 100 (in excellent health) [13]. Accordingly, by gathering these datasets, it is possible to estimate, with minimum delay, the incidence of ILI among Flusurvey participants, which has been previously shown to correlate with the incidence measured by sentinel-based surveillance at PHE [14].

Internet surveillance can complement traditional surveillance by measuring symptoms among a population with minimal time delay [15]. When this is done continually, it can improve the quality of the incidence data and help to inform policy decisions during routine seasonal influenza and pandemics [15]. The use of internet-based disease surveillance has increased rapidly in the past decade, including online ILI syndromic survey systems, such as the Europe-wide Influenzanet (of which Flusurvey is a member, working in synergy with other European platforms sharing data collection modality and results) [16], FluNearYou in the United States [17], and Australia's Flutracking [18]. Estimates of the relative incidence of influenza have also been inferred from search engine query data [19], detecting news reports from news sites aggregators [20], social media platforms such as Twitter [21], Wikipedia access logs [22], restaurant reservation and review logs [23], nonprescription pharmacy sales [24], and prediction markets [25]. Moreover, online crowd-sourced surveillance platforms have been similarly developed for other health conditions, including malaria [26], food-borne illness [27], and tick-borne diseases such as Lyme disease [28].

One criticism of online surveillance for influenza (and other syndromic surveillance mechanisms) is that because data are based on self-reporting of symptoms, the results collected are only representative of ILI, rather than virologically confirmed influenza [29]. Although previous years of ILI incidence from Flusurvey corresponded to RCGP influenza data, suggesting that Flusurvey does detect outbreaks of influenza, this has not been confirmed. Virological confirmation studies are required to assess how the measured ILI rates compare with the actual circulation of influenza in the population. Previously, two studies have sought to do this: one undertook a virological self-sampling from those calling the UK national telephone health helpline (NHS Direct) [30,31] and one completed virological self-sampling of GoViral participants in Massachusetts [32]. Similar self-testing or self-sampling studies have been undertaken for HIV/AIDS [33-35] and chlamydia, where the online cohort approach has been combined with a complete eSexual Health Clinic [36].

We designed a pilot study to assess whether it was feasible to detect laboratory-confirmed influenza from an online cohort. We monitored ILI through the Flusurvey platform, asking those reporting ILI to undertake a nasal swab for laboratory assessment of whether they were actually infected with an influenza virus, another respiratory virus, or unknown etiology (including viruses not tested for). This pilot study will help to assess the validity of online platforms for ILI surveillance and to confirm if the syndromic cohort surveillance approach is able to detect influenza infection. Moreover, it will additionally serve as a proof-of-concept study, showing that self-testing in the community can be successfully added to internet-based surveillance of ILI.

Methods

We recruited unvaccinated volunteer participants by including "Would you be interested in taking part in a virological self-swabbing study?" on Flusurvey's regular baseline recruitment survey [12]. Only unvaccinated participants were selected to offer increased chances of testing positive for influenza. Recruitment into the pilot was open from November 17, 2014 to December 17, 2014, with 1615 potential participants volunteering. As a feasibility study, limited by financial constraints, we sought to generate 100 swabs for testing. Based on experience from Flusurvey and the Flu Watch study [9], we estimated that approximately 48% of participants would experience a respiratory illness during periods of influenza circulation, 43% (3/7) of those would have an onset on the day or within the 2 days of reporting the illness (important for a high viral load to facilitate laboratory testing), and that 70% of those requested to return a swab would do so. Given these parameters a total of 700 individuals were recruited to take part in the self-swabbing pilot. Accordingly, we purposefully sampled 700 of 1615 eligible participants to include high-risk groups for influenza infection, including all those younger than 18 years and older than 65 years who volunteered or were volunteered by parents [37,38]. Random sampling of those aged between 18 and 64 years was then undertaken to reach a total of 700 participants. We sent these 700 participants a cover letter, an information sheet on the project, the virological swabbing kit, and sample transport materials (Multimedia Appendix 1). The virological swab kit contained a regular tip flocked swab plus 1 mL Universal Transport Medium (Sterilab, North Yorkshire, UK). For transportation of the sample to the laboratory, the pack contained a round mailing container (126 × 30 mm) with liner, a neutral screw cap for the mailing container, and a mailing box packaging system for containers up to 30 mm (Sarstedt, Numbrecht, Germany) to meet Category B UN3373 posting standards for viral materials. Also included were instructions for undertaking the swabbing (Multimedia Appendix 2), a link to video instructions [12], and identification labels for the samples (Multimedia Appendix 3). Participants were asked to store it safely until instructed to self-swab. Informed consent was obtained from pilot study participants when they selected that that they were willing to take part in the study online.

Participants were asked to report their symptoms online on a weekly basis. If their symptoms met the European Centre on Disease Prevention and Control definition of ILI, which is sudden onset of symptoms and at least one of four systemic symptoms (fever or feverishness, malaise, headache, myalgia) and at least one of three respiratory symptoms (cough, sore throat, shortness of breath) [39], and if the reported date of onset of symptoms was within 4 days of notification (to ensure virological testing during the acute phase of infection with expected higher viral loads [40]), then participants were asked to self-swab. Due to low circulation of influenza during the time period of the pilot study [14], the swabbing criteria were expanded on March 16, 2015, to include acute respiratory infection (ARI). (This comprises the sudden onset of symptoms and at least one of the following four respiratory symptoms: cough, sore throat, shortness of breath, and coryza [39].)

Nasal swabbing was chosen because it has been shown to be effective for influenza testing by patients [41] and offers a greater viral load than saliva collection [32]. Although nasopharyngeal swabbing may yield a slightly greater viral load for confirmation with polymerase chain reaction (PCR) tests [42], we believed it would be difficult to self-administer nasopharyngeal swabs of consistent quality and therefore flocked nasal swabs provided a suitable alternative [43]. Participants were provided with both written and video instructions as to how to administer the self-swab (Multimedia Appendix 2).

Participants were asked to send their samples (Multimedia Appendix 2) to the PHE Laboratory, Bristol, for respiratory virus PCR testing. The samples were assayed using PHE Bristol in-house–validated PCR panels including targets for common respiratory infections, including influenza A, influenza B, respiratory syncytial virus (A and B) human metapneumovirus, parainfluenza virus (1, 2, and 3), adenovirus, and rhinovirus. This assay was chosen based on its confidence in detecting circulating strains of influenza. The PHE Laboratory used a generic influenza A assay that is assessed *in silico* against common circulating strains (primer and probe matching) and then tested in practice using a proficiency panel constructed by the respiratory virus unit from PHE, plus other external quality assurance schemes. Results were returned to researchers at the London School of Hygiene and Tropical Medicine and sent to participants via email. As this was a pilot research study, rather than a clinical test, participants were informed before self-sampling that their results would not be available in real time.

It was important to ascertain whether samples with negative results through the PCR multiplex testing were truly negative or whether the test had not been administered properly, providing a false negative. Second-stage testing was undertaken at London Centre for Nanotechnology, University College London. These samples were quantified for human nuclear and mitochondrial DNA by TaqMan PCR for mammalian glyceraldehyde 3-phosphate dehydrogenase (GAPDH) DNA in the swab samples [44]. Positive controls for the reaction were either HeLa cell DNA (NEB, Hitchin, UK) or human placental DNA (Sigma, Dorset, UK). Cycling conditions were 95°C for 15 seconds and annealing/extension at 60°C for 1 minute after an initial denaturation of 10 minutes.

Finally, we asked those participants to complete a short evaluation form online to ascertain how easy they found the process and its viability for future (Multimedia Appendix 4).

Ethics approval for undertaking this study was obtained from Observational and Interventions Research Ethics Committee at London School of Hygiene and Tropical Medicine (ref: 5530-03).

Results

Samples were received at PHE Laboratory, Bristol, from January 1, 2015 to April 7, 2015. In total, 66 participants from the pilot group (of the 100 originally estimated based on sample size

calculations) met the symptom and timing of onset criteria and were asked to self-swab. A total of 51 swab samples (77%) were received at PHE Laboratory. An additional three samples were received, although they were not requested, and therefore they were not included in the analysis. Multiplex PCR testing results are presented in Table 1 and Multimedia Appendix 5.

The second-stage testing for the presence of human DNA produced the results presented in Table 2 and Multimedia Appendix 6.

These findings show all samples contained human DNA and are consistent with the correct use of the swab, corroborating earlier successful experiences of self-swabbing in a home setting [9,32,41]. This validated the feasibility and the process used as well as the respiratory virus detection results.

Reflecting similar demographic trends from the broader Flusurvey project [45], including from the 2014-2015 cohort from which these participants were selected [46], the swabs received at the laboratory for virological testing were not representative of the UK population. Of the 51 results received, 36 were received from female participants and 15 were from male participants. The age of participants ranged from 4 to 91 years with a mean age of 41 (SD 19) years.

As part of the routine information collected by Flusurvey, participants are asked to score how they were feeling each week [13]. Over the course of the 2014-2015 influenza season, the mean reported score for all Flusurvey users (N=6102) was 82.9. This is self-reported on a scale from 0 to 100. For the 66 self-swab participants, the mean score for the weeks they reported ILI (or ILI and ARI after March 16, 2015) was 63.9 (SD 23.9). From those who tested positive for ILI, the range of scores was 30 to 100 and the mean was 72.8 (SD 16.9). Finally, for those who tested positive for influenza, the range was 60 to 85 and the mean was 72 (SD 9). Although there is a great risk of overinterpretation with a small sample size, and a risk of bias by characteristics of people self-swabbing, these results do not suggest a positive relationship between incidence of ILI and (self-reported) severity of symptoms through the health score.

Completed evaluation forms were received from 21 participants. Of these, 20 participants suggested that undertaking the swab was easy or very easy, although one stated it was “unpleasant.” In addition, 13 participants indicated that if they were to have ILI symptoms in the future, they would prefer to undertake a self-swab at home to diagnose symptoms, five participants preferred to treat symptoms at home without a swab, and three participants were undecided about their future use of swabbing at home. All those who completed the evaluation form found the written instructions helpful, and 13 participants found the video instructions useful, with the remaining eight participants not watching it.

Table 1. Yield of influenza-like illness positive tests from samples tested (N=51) using multiplex quantitative PCR..

Virus	Positive tests, n (%)
Influenza A (H3N2)	1
Influenza B	4
Human metapneumovirus[a]	2
Rhinovirus[a]	11
Inhibitory samples[b]	2
Total viral yield	18 (35)
Total influenza yield	5 (10)

[a]One sample tested positive for both human metapneumovirus and rhinovirus. We have included both of these infections separately in this table, but this reflects one dual infection.

[b]Two samples contained inhibitory substances that hindered amplification of sample control markers. No conclusions can therefore be drawn from these samples, positive or negative.

Table 2. Detection of mammalian DNA using Taqman quantitative polymerase chain reaction.

Group	Samples, n	GAPDH level (Ct[a]), mean (SD)
No virus detected	20	28.4 (2.9)
Virus detected	11	28.7 (2.5)
Low virus detected	5	28.2 (4.7)
Inhibitors present	2	26.1 (3.7)
Not tested for virus	3	23.7 (1.6)

[a]Ct: cycle threshold. A Ct value <40 represents positive detection of human DNA.

Discussion

This study has shown that, as a proof of concept, it is possible to successfully apply an at-home self-swabbing methodology to an internet-based cohort and that this can detect both influenza and other causes of ILI by collecting viral samples of suitable quality for PCR multiplex testing. This replicated findings of similar studies conducted for self-swabbing from an online cohort through the GoViral Platform in the United States, through phone-based surveillance via NHS Direct, and in community self-swabbing through Flu Watch [9,31,32], extending these to assess feasibility among a crowd-sourced platform. It was estimated that in the 2014-2015 influenza season, there were low to moderate levels of influenza activity with the predominant strain influenza A (H3N2) [14] present for the majority of the season (the majority were antigenically similar to the A/Texas/50/2012 H3N2 Northern Hemisphere strain), and the appearance of influenza B during the last months of the season (the majority of these belonged to the B/Yamagata 16/88 lineage) [14]. This pilot study did not generate a large yield of ILI virus-positive results, nor was it powered to, yet there were similar trends between this pilot and PHE Laboratory–confirmed cases for the same time period and in particular from the RCGP sentinel swabbing scheme, to which this would be most similar for detecting influenza in a community cohort [14].

Participants found the self-swabbing easy to undertake and several individuals in the pilot indicated that they would be interested in using a similar self-swab in the future if they wanted laboratory confirmation of ILI infection. This was a research project, therefore testing of the samples was not undertaken in real time and results were not returned to participants for a number of weeks, by which time their symptoms would have likely subsided, meaning that the results of the virological test would not have affected their behavior. However, if self-swabbing were to be conducted and results returned in real time, the results may affect patient behavior, either through visiting a health professional, taking medication, or changing their daily routine to limit potential viral spread. Moreover, virological confirmation of a viral infection may potentially reduce antibiotic prescription due to misdiagnosis.

Usability of Findings

Yet, we do not wish to suggest that this methodology should replace the efficient RCGP sentinel swabbing system; our approach may not be practical for routine influenza surveillance owing to the cost and logistics of distributing kits. However, self-swabbing of an internet-based cohort may prove useful for ad hoc surveys, such as in the emerging stages of a pandemic to understand community transmission or as a supplementary tool to otherwise established surveillance mechanisms. Our study can also contribute to demonstrating the feasibility of both an online cohort approach to surveillance and self-swabbing at home for other health conditions, such as sexually transmitted infections or gastroenteritis for which there might be privacy reasons for patients seeking to test themselves at home.

Although this pilot study cannot make conclusive remarks about the validity of online influenza surveillance, it has shown that, as a proof of concept, it is possible to detect an influenza virus and other ILI from a cohort of online participants. Accordingly, this can be replicated at larger scale for greater verification of online crowd-sourced disease surveillance mechanisms. Moreover, if a self-swabbing study were to be repeated with a greater number of participants and samples, from a more representative demographic sample, it may be possible to build these into a strong analytical model for estimating the burden of influenza.

Limitations

Due to delays in procuring the necessary materials, delays in obtaining ethical approval, postal delays, and closures due to Christmas, the self-swab kits were not distributed to participants until the first week of January 2015. Retrospectively, it can be seen that the peak of the influenza season in the UK during the 2014-2015 season was week 52 (December 22-28, 2014) [14]. As such, the study did not take place during the peak influenza season in the UK and there were not high levels of influenza circulating during the pilot period. Accordingly, we did not have a large sample group of swabs nor did we obtain the predicted 100 swabs. Despite the methodological change to include ARI on March 16, 2015 [39], this still failed to collect the original requirement of 100 samples. This trend matched PHE's microbiological surveillance for the same time period, with a similarly low yield of influenza-positive samples [14]. However, the pilot study did coincide with the later peak of influenza B [14], and hence detections of influenza B.

If this project were to be undertaken again, swab kits should be distributed to participants before the start of the influenza season to mitigate the uncertainty of predicting when the virus may arrive and/or be at its peak. Alternatively, kits should be sent out, or be collectable locally, after notification of relevant symptoms. Although this delay may impact the viral load collected, it would prove more cost effective in a health care setting.

The participants who sent samples to the laboratory for testing were not a representative sample demographically, featuring predominantly women aged between 18 and 64 years in Southeast England. This reflects Flusurvey and Influenzanet participants more generally and does not represent a random sample of the UK/European populations [47]. Any future self-swabbing studies carried out on an internet-based cohort could broaden the demographics of the sample to increase the study's applicability. This could include greater recruitment drives among underrepresented groups or more purposeful sampling to take the bias of the wider Flusurvey group into account.

A further limitation was that not all respiratory viruses were tested for (eg, coronavirus and enterovirus). However, because this was a proof-of-concept study for Flusurvey, a crowd-sourcing platform for influenza surveillance, the focus remained on assessing influenza and other respiratory infections were considered to be supplementary. This may account for the individuals who reported symptoms, but whose swabs tested negative for influenza.

Conclusion

This pilot study has shown that it is feasible for individuals to conduct self-swabbing for ILI/ARI in their own home at relatively low cost. Those selected to participate were able to successfully collect samples and the biological material gathered was sufficient for influenza and other viruses to be detected in the laboratory. This allows us to conclude that, as a proof of concept, it is possible to use home swabbing for detection of influenza at the community level. Due to the small sample size, conclusive statements about how effective the Flusurvey algorithms may be in comparison to other forms of community-based surveillance cannot be made, yet it still validates the conceptual approach used for online symptomatic surveillance methodology. However, there remain concerns about the accuracy of such a system and further research would be needed to repeat a similar experiment with a greater number of participants to provide a suitable sample size to make any broader assumptions about the accuracy of online influenza detection systems.

Acknowledgments

This research was funded by an i-sense exploratory project, as part of a larger Engineering and Physical Sciences Research Council UK (EPSRC) project EP/K031953/1.

We thank the wider i-sense project, of which this was one part, and Rumi Chunara and the GoViral study for their advice on development of the study. We also thank all the anonymous pilot study participants and Flusurvey users who contributed to this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Information sheet for pilot participants.

[PDF File (Adobe PDF File), 202KB - jmir_v20i3e71_app1.pdf]

Multimedia Appendix 2

Written instructions for self-swabbing.

[PDF File (Adobe PDF File), 103KB - jmir_v20i3e71_app2.pdf]

Multimedia Appendix 3

Swab sample labels for laboratory testing.

[PDF File (Adobe PDF File), 62KB - jmir_v20i3e71_app3.pdf]

Multimedia Appendix 4

Pilot evaluation form.

[PDF File (Adobe PDF File), 278KB - jmir_v20i3e71_app4.pdf]

Multimedia Appendix 5

Anonymized dataset.

[XLSX File (Microsoft Excel File), 35KB - jmir_v20i3e71_app5.xlsx]

Multimedia Appendix 6

Additional dataset from secondary testing.

[XLS File (Microsoft Excel File), 54KB - jmir_v20i3e71_app6.xls]

References

1. World Health Organization. Influenza (seasonal) factsheet URL: http://www.who.int/mediacentre/factsheets/fs211/en/ [accessed 2018-02-07] [WebCite Cache ID 6x3ElIzNy]
2. Keech M, Beardsworth P. The impact of influenza on working days lost: a review of the literature. Pharmacoeconomics 2008;26(11):911-924. [Medline: 18850761]

3. Public Health England. 2014 Oct 06. Public Health England and the NHS prepare for unpredictable flu season URL: https://www.gov.uk/government/news/public-health-england-and-the-nhs-prepare-for-unpredictable-flu-season [accessed 2018-02-07] [WebCite Cache ID 6x3F5EkdJ]
4. Pitman RJ, White LJ, Sculpher M. Estimating the clinical impact of introducing paediatric influenza vaccination in England and Wales. Vaccine 2012 Feb 01;30(6):1208-1224 [FREE Full text] [doi: 10.1016/j.vaccine.2011.11.106] [Medline: 22178725]
5. Pockett RD, Watkins J, McEwan P, Meier G. Burden of illness in UK subjects with reported respiratory infections vaccinated or unvaccinated against influenza: a retrospective observational study. PLoS One 2015;10(8):e0134928 [FREE Full text] [doi: 10.1371/journal.pone.0134928] [Medline: 26287532]
6. Smith RD, Keogh-Brown MR, Barnett T. Estimating the economic impact of pandemic influenza: an application of the computable general equilibrium model to the U.K. Soc Sci Med 2011 Jul;73(2):235-244. [doi: 10.1016/j.socscimed.2011.05.025] [Medline: 21708419]
7. Briand S, Mounts A, Chamberland M. Challenges of global surveillance during an influenza pandemic. Public Health 2011 May;125(5):247-256. [doi: 10.1016/j.puhe.2010.12.007] [Medline: 21524774]
8. Porteous T, Ryan M, Bond CM, Hannaford P. Preferences for self-care or professional advice for minor illness: a discrete choice experiment. Br J Gen Pract 2006 Dec;56(533):911-917 [FREE Full text] [Medline: 17132378]
9. Hayward AC, Fragaszy EB, Bermingham A, Wang L, Copas A, Edmunds WJ, Flu Watch Group. Comparative community burden and severity of seasonal and pandemic influenza: results of the Flu Watch cohort study. Lancet Respir Med 2014 Jun;2(6):445-454 [FREE Full text] [doi: 10.1016/S2213-2600(14)70034-7] [Medline: 24717637]
10. Mandl KD, Overhage JM, Wagner MM, Lober WB, Sebastiani P, Mostashari F, et al. Implementing syndromic surveillance: a practical guide informed by the early experience. J Am Med Inform Assoc 2004;11(2):141-150 [FREE Full text] [doi: 10.1197/jamia.M1356] [Medline: 14633933]
11. Public Health England. 2014 Jan 07. Sources of UK flu data: influenza surveillance in the UK URL: https://www.gov.uk/guidance/sources-of-uk-flu-data-influenza-surveillance-in-the-uk [accessed 2018-02-07] [WebCite Cache ID 6x3HM5qgg]
12. Influenzanet. Flusurvey URL: https://flusurvey.org.uk/ [accessed 2018-02-07] [WebCite Cache ID 6x3HQHKVG]
13. Adler AJ, Eames KT, Camacho A, Edmunds WJ. Health scores in Flusurvey participants: findings from the 2012–13 influenza season. Lancet 2013 Nov;382:S22. [doi: 10.1016/S0140-6736(13)62447-2]
14. Public Health England. Surveillance of Influenza and Other Respiratory Viruses in the United Kingdom: Winter 2014 to 2015. London: Public Health England; 2015 Mar. URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/429617/Annualreport_March2015_ver4.pdf [accessed 2018-02-07] [WebCite Cache ID 6x3I09NDL]
15. Brooks-Pollock E, Tilston N, Edmunds W, Eames KT. Using an online survey of healthcare-seeking behaviour to estimate the magnitude and severity of the 2009 H1N1v influenza epidemic in England. BMC Infect Dis 2011 Mar 16;11:68 [FREE Full text] [doi: 10.1186/1471-2334-11-68] [Medline: 21410965]
16. Paolotti D, Carnahan A, Colizza V, Eames K, Edmunds J, Gomes G, et al. Web-based participatory surveillance of infectious diseases: the Influenzanet participatory surveillance experience. Clin Microbiol Infec 2014 Jan;20(1):17-21. [doi: 10.1111/1469-0691.12477]
17. Smolinski MS, Crawley AW, Baltrusaitis K, Chunara R, Olsen JM, Wójcik O, et al. Flu near you: crowdsourced symptom reporting spanning 2 influenza seasons. Am J Public Health 2015 Oct;105(10):2124-2130. [doi: 10.2105/AJPH.2015.302696] [Medline: 26270299]
18. Dalton C, Durrheim D, Fejsa J, Francis L, Carlson S, d'Espaignet ET, et al. Flutracking: a weekly Australian community online survey of influenza-like illness in 2006, 2007 and 2008. Commun Dis Intell Q Rep 2009 Sep;33(3):316-322 [FREE Full text] [Medline: 20043602]
19. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
20. Brownstein JS, Freifeld CC, Reis BY, Mandl KD. Surveillance Sans Frontières: Internet-based emerging infectious disease intelligence and the HealthMap project. PLoS Med 2008 Jul 08;5(7):e151 [FREE Full text] [doi: 10.1371/journal.pmed.0050151] [Medline: 18613747]
21. Aramaki E, Sachiko M, Mizuki M. Twitter catches the flu: detecting influenza epidemics using Twitter. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2011 Presented at: 2011 Conference on Empirical Methods in Natural Language Processing; Jul 27–31, 2011; Edinburgh, Scotland p. 1568-1576.
22. McIver DJ, Brownstein JS. Wikipedia usage estimates prevalence of influenza-like illness in the United States in near real-time. PLoS Comput Biol 2014 Apr;10(4):e1003581 [FREE Full text] [doi: 10.1371/journal.pcbi.1003581] [Medline: 24743682]
23. Nsoesie EO, Buckeridge DL, Brownstein JS. Guess who's not coming to dinner? Evaluating online restaurant reservations for disease surveillance. J Med Internet Res 2014 Jan 22;16(1):e22 [FREE Full text] [doi: 10.2196/jmir.2998] [Medline: 24451921]
24. Das D, Metzger K, Heffernan R, Balter S, Weiss D, Mostashari F, New York City Department of Health and Mental Hygiene. Monitoring over-the-counter medication sales for early detection of disease outbreaks--New York City. MMWR Suppl 2005 Aug 26;54:41-46. [Medline: 16177692]

25. Polgreen PM, Nelson FD, Neumann GR. Use of prediction markets to forecast infectious disease activity. Clin Infect Dis 2007 Jan 15;44(2):272-279. [doi: 10.1086/510427] [Medline: 17173231]
26. Chunara R, Chhaya V, Bane S, Mekaru SR, Chan EH, Freifeld CC, et al. Online reporting for malaria surveillance using micro-monetary incentives, in urban India 2010-2011. Malar J 2012 Feb 13;11(1):43 [FREE Full text] [doi: 10.1186/1475-2875-11-43] [Medline: 22330227]
27. Quade P, Nsoesie E. A platform for crowdsourced foodborne illness surveillance: description of users and reports. JMIR Public Health Surveill 2017 Jul 05;3(3):e42 [FREE Full text] [doi: 10.2196/publichealth.7076] [Medline: 28679492]
28. Hines D, Sibbald SL. Public Health Agency of Canada. 2015. Citizen science exploring its application as a tool for prodromic surveillance of vector-borne disease.Canada Communicable Disease Report URL: https://www.canada.ca/en/public-health/services/reports-publications/canada-communicable-disease-report-ccdr/monthly-issue/2015-41/ccdr-volume-41-03-march-5-2015/ccdr-volume-41-03-march-5-2015-3.html [accessed 2018-02-07] [WebCite Cache ID 6x3JVfLBR]
29. Adler AJ, Eames KT, Funk S, Edmunds WJ. Incidence and risk factors for influenza-like-illness in the UK: online surveillance using Flusurvey. BMC Infect Dis 2014 May 01;14:232 [FREE Full text] [doi: 10.1186/1471-2334-14-232] [Medline: 24885043]
30. Cooper DL, Smith GE, Chinemana F, Joseph C, Loveridge P, Sebastionpillai P, et al. Linking syndromic surveillance with virological self-sampling. Epidemiol Infect 2008 Feb;136(2):222-224. [doi: 10.1017/S0950268807008412] [Medline: 17394678]
31. Elliot AJ, Bermingham A, Charlett A, Lackenby A, Ellis J, Sadler C, et al. Self-sampling for community respiratory illness: a new tool for national virological surveillance. Euro Surveill 2015 Mar 12;20(10):21058 [FREE Full text] [Medline: 25788252]
32. Goff J, Rowe A, Brownstein JS, Chunara R. Surveillance of acute respiratory infections using community-submitted symptoms and specimens for molecular diagnostic testing. PLoS Curr 2015 May 27;7:27 [FREE Full text] [doi: 10.1371/currents.outbreaks.0371243baa7f3810ba1279e30b96d3b6] [Medline: 26075141]
33. MacPherson P, Lalloo D, Webb E, Maheswaran H, Choko A, Makombe S, et al. Effect of optional home initiation of HIV care following HIV self-testing on antiretroviral therapy initiation among adults in Malawi: a randomized clinical trial. JAMA 2014;312(4):372-379 [FREE Full text] [doi: 10.1001/jama.2014.6493] [Medline: 25038356]
34. Ng OT, Chow AL, Lee VJ, Chen MIC, Win MK, Tan HH, et al. Accuracy and user-acceptability of HIV self-testing using an oral fluid-based HIV rapid test. PLoS One 2012;7(9):e45168 [FREE Full text] [doi: 10.1371/journal.pone.0045168] [Medline: 23028822]
35. Aghaizu A, Wayal S, Nardone A, Parsons V, Copas A, Mercey D, et al. Sexual behaviours, HIV testing, and the proportion of men at risk of transmitting and acquiring HIV in London, UK, 2000-13: a serial cross-sectional study. Lancet HIV 2016 Sep;3(9):e431-e440. [doi: 10.1016/S2352-3018(16)30037-6] [Medline: 27562744]
36. Estcourt C, Gibbs J, Sutcliffe L, Gkatzidou V, Tickle L, Hone K, et al. The eSexual Health Clinic system for management, prevention, and control of sexually transmitted infections: exploratory studies in people testing for Chlamydia trachomatis. Lancet Publ Health 2017 Apr;2(4):e182-e190 [FREE Full text] [doi: 10.1016/S2468-2667(17)30034-8] [Medline: 29253450]
37. Cromer D, van Hoek AJ, Jit M, Edmunds WJ, Fleming D, Miller E. The burden of influenza in England by age and clinical risk group: a statistical analysis to inform vaccine policy. J Infect 2014 Apr;68(4):363-371 [FREE Full text] [doi: 10.1016/j.jinf.2013.11.013] [Medline: 24291062]
38. Gross PA, Hermogenes AW, Sacks HS, Lau J, Levandowski RA. The efficacy of influenza vaccine in elderly persons. A meta-analysis and review of the literature. Ann Intern Med 1995 Oct 01;123(7):518-527. [Medline: 7661497]
39. European Center for Disease Control. Influenza case definitions URL: http://ecdc.europa.eu/en/healthtopics/influenza/surveillance/Pages/influenza_case_definitions.aspx [accessed 2018-02-07] [WebCite Cache ID 6x3KEUIWS]
40. Druce J, Garcia K, Tran T, Papadakis G, Birch C. Evaluation of swabs, transport media, and specimen transport conditions for optimal detection of viruses by PCR. J Clin Microbiol 2012 Mar;50(3):1064-1065 [FREE Full text] [doi: 10.1128/JCM.06551-11] [Medline: 22205810]
41. Dhiman N, Miller RM, Finley JL, Sztajnkrycer MD, Nestler DM, Boggust AJ, et al. Effectiveness of patient-collected swabs for influenza testing. Mayo Clin Proc 2012 Jun;87(6):548-554 [FREE Full text] [doi: 10.1016/j.mayocp.2012.02.011] [Medline: 22551906]
42. Heikkinen T, Salmi AA, Ruuskanen O. Comparative study of nasopharyngeal aspirate and nasal swab specimens for detection of influenza. BMJ 2001 Jan 20;322(7279):138 [FREE Full text] [Medline: 11159569]
43. Akmatov MK, Gatzemeier A, Schughart K, Pessler F. Equivalence of self- and staff-collected nasal swabs for the detection of viral respiratory pathogens. PLoS One 2012;7(11):e48508 [FREE Full text] [doi: 10.1371/journal.pone.0048508] [Medline: 23155387]
44. Gray ER, Garson JA, Breuer J, Edwards S, Kellam P, Pillay D, et al. No evidence of XMRV or related retroviruses in a London HIV-1-positive patient cohort. PLoS ONE 2011 Mar 23;6(3):e18096. [doi: 10.1371/journal.pone.0018096]
45. Tilston NL, Eames KT, Paolotti D, Ealden T, Edmunds WJ. Internet-based surveillance of Influenza-like-illness in the UK during the 2009 H1N1 influenza pandemic. BMC Public Health 2010 Oct 27;10:650 [FREE Full text] [doi: 10.1186/1471-2458-10-650] [Medline: 20979640]

46. Peppa M, John EW, Funk S. Disease severity determines health-seeking behaviour amongst individuals with influenza-like illness in an internet-based cohort. BMC Infect Dis 2017 Dec 31;17(1):238 [FREE Full text] [doi: 10.1186/s12879-017-2337-5] [Medline: 28359335]
47. Cantarelli P, Debin M, Turbelin C, Poletto C, Blanchon T, Falchi A, et al. The representativeness of a European multi-center network for influenza-like-illness participatory surveillance. BMC Public Health 2014 Sep 20;14:984 [FREE Full text] [doi: 10.1186/1471-2458-14-984] [Medline: 25240865]

Abbreviations

ARI: acute respiratory infection
GP: general practitioner
ILI: influenza-like illness
PCR: polymerase chain reaction
PHE: Public Health England
RCGP: Royal College of General Practitioners

Edited by G Eysenbach; submitted 29.09.17; peer-reviewed by H Miller, A Hayward; comments to author 01.11.17; revised version received 14.11.17; accepted 15.11.17; published 01.03.18.

Please cite as:
Wenham C, Gray ER, Keane CE, Donati M, Paolotti D, Pebody R, Fragaszy E, McKendry RA, Edmunds WJ
Self-Swabbing for Virological Confirmation of Influenza-Like Illness Among an Internet-Based Cohort in the UK During the 2014-2015 Flu Season: Pilot Study
J Med Internet Res 2018;20(3):e71
URL: http://www.jmir.org/2018/3/e71/
doi:10.2196/jmir.9084
PMID:29496658

Original Paper

Evolution of Electronic Cigarette Brands From 2013-2014 to 2016-2017: Analysis of Brand Websites

Greta Hsu[1], PhD; Jessica Y Sun[2], BA; Shu-Hong Zhu[2,3], PhD

[1]Graduate School of Management, University of California Davis, Davis, CA, United States

[2]Moores Cancer Center, University of California San Diego, La Jolla, CA, United States

[3]Department of Family Medicine and Public Health, University of California San Diego, La Jolla, CA, United States

Corresponding Author:
Shu-Hong Zhu, PhD
Department of Family Medicine and Public Health
University of California San Diego
9500 Gilman Drive
MC 0905
La Jolla, CA, 92093
United States
Phone: 1 8583001056
Email: szhu@ucsd.edu

Abstract

Background: The electronic cigarette (e-cigarette) industry has grown in size and organizational complexity in recent years, most notably with the entry of major tobacco companies in 2012 and the proliferation of vape shops. Many brands maintain retail websites that present e-cigarette marketing claims and sell directly to consumers. Understanding of the evolving composition of different types of e-cigarette brand websites is currently underdeveloped.

Objective: This paper presents how e-cigarette brand websites surveyed in 2013-2014 evolved by 2016-2017, and how the websites run by different types of e-cigarette producers currently differ.

Methods: In 2016-2017, we revisited 466 e-cigarette brand websites surveyed in 2013-2014, 288 of which were extant, and identified 145 new English-language websites. We compared product designs, marketing claims, and age-based warnings presented by types of e-cigarette producers: major tobacco companies, independent vape shops, and independent internet-only companies.

Results: Among the 433 websites examined in 2016-2017, 12 were owned by major tobacco companies, 162 operated a physical vape shop, and 259 were internet-only operations. Closed-system product designs were sold by 83% (10/12) of tobacco-owned brands. In comparison, 29.0% (47/162, *P*<.001) of vape shop and 55.2% (143/259, *P*=.06) of internet-only brands sold closed-system designs. Compared with vape shop and internet-only brands, tobacco-owned brands offered a smaller set of product models (*P* values <.001) and a narrower range of flavors (*P* values <.01), with greater emphasis on the traditional combustible cigarette flavors of tobacco and menthol (*P* values <.001). Tobacco-owned brands also offered a narrower range of nicotine options than the vape shops (*P*=.002) and were less likely to offer nicotine-free e-liquid compared with internet-only and vape shop brands (*P* values <.001). Finally, 83% (10/12) of tobacco-owned brand websites featured age verification pop-up windows. In comparison, only 50.2% (130/259) of internet-only brands (*P*=.01) and 60.5% (98/162) of vape shop brands (*P*=.06) featured age verification windows. Websites surveyed in both 2013-2014 and 2016-2017 became more likely to sell open-system mods (*P*<.001) and sold an increased number of product models (*P*<.001), flavors (*P*<.001), and nicotine options (*P*<.001). Prevalence of several types of claims decreased significantly, including indirect claims regarding smoking cessation (*P*<.001), claims regarding e-cigarettes as healthier (*P*<.001), less expensive (*P*<.001), and usable in more places (*P*<.001) compared with combustible cigarettes.

Conclusions: The number of e-cigarette brands has not appeared to increase since 2014, even as website messaging evolved, with brands owned by tobacco companies and vape shops pulling in opposite directions. Brands owned by tobacco companies offered a limited range of e-cigarette products, whereas brands owned by vape shops emphasized a panoply of flavor and nicotine options. Furthermore, the Food and Drug Administration's regulatory action may influence the types of e-cigarette products offered and the market shares of various companies.

(J Med Internet Res 2018;20(3):e80) doi:10.2196/jmir.8550

KEYWORDS

electronic cigarettes; vaping; nicotine; longitudinal studies; internet; market research

Introduction

Background

Since its introduction in the US market in 2007, the electronic nicotine delivery system (ENDS) industry has grown to an estimated US $3.5 billion market [1] with 3 types of electronic cigarette (e-cigarette) products: cigalikes, eGos (or pen-style e-cigarettes), and mods. Cigalikes came on the market first, with slim cylindrical closed-system designs that use prefilled cartridges to maximize the ease of use. eGos and mods are advanced open-system designs that allow users to fill their own e-liquid solution, and they often have adjustable e-liquid heating temperatures, allowing a customized nicotine yield and puff volume [2-5]. eGos have larger cylindrical shapes and stronger batteries than cigalikes, whereas mods are the most customizable and come in a wide range of shapes and sizes.

As the e-cigarette market's product landscape has evolved, so has its organizational composition. Major tobacco companies such as Lorillard, Altria Group, and Reynolds American began entering the e-cigarette industry in 2012 and have increasingly dominated its market share [6]. Vape shops—independent retail shops that specialize in ENDS products—also appear to have proliferated at a rapid pace [7-9]. Many of these different types of brands maintain retail websites that present marketing claims about e-cigarettes and sell ENDS products directly to consumers. A large proportion of e-cigarette sales are conducted through Web-based channels, which due to the industry's young and historically unregulated status have not been well tracked [10].

Prior studies conducted by public health researchers reveal important patterns in the advancement of marketing claims. For example, e-cigarette websites often display claims about health or smoking cessation benefits of e-cigarette [11,12], and older e-cigarette brands are more likely to advance claims regarding harm reduction and smoking cessation relative to newer brands [13]. Studies have also found differences in products sold across different types of brands. In 2015, researchers found that major tobacco companies were likely to offer e-cigarette products with closed-system designs and more limited flavors than independents [14].

One large-scale survey documented products sold and claims advanced by 466 e-cigarette websites from December 2013 to January 2014 [13]. We conduct a follow-up to that survey for the period surrounding the finalization of the Food and Drug Administration's (FDA's) Deeming Rule in August 2016, which extended the FDA's regulatory authority under the Family Smoking Prevention and Tobacco Control Act to the ENDS industry [15]. The Deeming Rule has generated a host of varied reactions, ranging from praise for reining in an unregulated product increasingly popular among youths to concern that strict regulatory oversight will limit smokers' options for an increasingly popular method for smoking cessation [16]. Given that the FDA recently announced plans to review and revise its regulatory oversight rules [17], gaining a better understanding of the products, marketing claims, and age-based protections offered by different types of brands may help regulators better anticipate the impact of regulatory oversight on this young industry's evolution. Differences along these dimensions are potentially important because they relate to the kinds of products e-cigarette users have access to, the information regarding e-cigarettes that potential and current e-cigarette users are presented with, and the ability of youths to gain access to e-cigarettes.

Study Question

This study examined all the websites from the 2013-2014 study to determine how many continued to operate in 2016-2017. In addition, we conducted another broad search for e-cigarette brands sold online in 2016-2017. We used the same methodology as the 2013-2014 survey, with particular attention to the ownership of these brands—whether they were owned by major tobacco companies, independent vape shops, or independent internet-only retail companies.

Two main issues were considered. First, did the product designs sold, claims advanced, and age-based protections offered by major tobacco-owned brands, internet-only, and vape shop brands differ during the more recent period? Second, how have brands sampled back in 2013-2014 evolved? For example, did brands change their proclivity to advance claims regarding harm reduction and smoking cessation? Prior studies have conducted surveys of e-cigarette brand websites at single points in time [11,12,14,18], and one study sampled the industry for 2 consecutive years [19]. However, we are not aware of studies tracking the same set of websites over time. General patterns in these changes may provide some indication of how industry incumbents will continue to evolve as this young industry develops. After documenting e-cigarette brands sold on the internet, we consider the potential public health and regulatory implications of our findings. Although we cannot definitively determine how regulatory oversight will shape this young market, our findings suggest potential issues that regulators should consider going forward.

Methods

Internet Search

To create our dataset, we first searched the list of websites surveyed in 2013-2014 [13] to identify those still conducting online retail operations as of July 2016. To this dataset, we added websites based on a new internet search of e-cigarette brands, conducted also in July 2016. This search mirrored the structure of the earlier surveys. Using 3 search engines (Google, Yahoo, and Bing), we searched for e-cigarette brands using the following keywords: e-cigarette, e cigarette, e-cig, e cig, ecig, ecigs, electronic cigarette, electronic cig, electronic nicotine delivery system, vape, vaper, and vaping. A website was included as an e-cigarette brand website if it sold e-cigarette hardware and identified at least one hardware or e-liquid product as its own. We included all e-cigarette brand websites listed on the first 30 pages of each search, excluding the following:

non-English websites; websites that did not sell products directly to the general public (eg, product review sites, manufacturer sites); websites selling cannabis-only products; and resale sites (eg, Amazon, eBay).

Brands and Models

Through a 2016 US Department of Health and Human Services industry report, organizational information listed on brand websites, and publicly available information on the internet (eg, company press releases), we identified brands owned by major tobacco companies [6]. Among the remaining brands, we distinguished between those that operated their own physical vape shop versus internet-only operations. Brands were treated as operating a physical vape shop if they ran a retail brick-and-mortar store selling ENDS products. Brands that only operated at mall kiosks and at locations that only allow for pickup of online sales were not considered as vape shops.

From August 2016 to February 2017, a project manager and 5 trained research assistants coded the websites for age-based access restrictions, product characteristics, product claims, e-liquid flavors, ingredients, and nicotine strengths. Coding was based on the 2013-2014 survey codebook. To ensure consistency with the 2013-2014 coding, the project manager overseeing the earlier survey provided training and extensive consultation to the project manager for the 2016-2017 survey. When training the research assistants, an initial set of websites were coded by each of the research assistants to identify and resolve any discrepancies in coding approaches, with extensive feedback given to ensure consistency in approaches. To further ensure consistency, regular checks of the website codes were conducted by the project manager throughout the coding process.

Each e-cigarette sold through a brand website was coded as cigalike, eGo, or mod. Every distinct e-cigarette model sold, including those of competitors' brands sold on a brand's website, were coded. The e-cigarettes that only varied in color or flavor of e-liquid were not counted as separate models.

Claims and Disclaimers About Electronic Cigarettes

Research assistants reviewed entire websites to determine the presence or absence of several types of claims and disclaimers. The following smoking cessation–related statements were coded: (1) a direct claim of e-cigarettes as an effective quitting aid; (2) an indirect claim (eg, a featured customer testimonial) of e-cigarettes as an effective quitting aid; and (3) a disclaimer that e-cigarettes are not approved as smoking cessation devices. We also coded whether websites claimed e-cigarettes are healthier/safer in comparison with combustible cigarettes. Additionally, we coded for claims regarding social benefits, including that e-cigarettes (1) are less expensive, (2) can be smoked in more places, (3) are cleaner or less messy/smelly, and (4) are more socially accepted when compared with combustible cigarettes. Age-related disclaimers were also coded, including a disclaimer that e-cigarettes are not intended for youths/minors and the presence of a pop-up window that asks website visitors to self-report either their age or whether they meet a minimum age threshold.

Flavors

The research assistants recorded whether the websites sold e-liquid (in either prefilled or liquid bottle formats), and they also recorded every distinct flavor of e-cigarette sold by each website. Distinct flavors were indicated by distinct linguistic labels for flavor (eg, "Cinnamon" and "Red Hot Cinnamon" were treated as distinct flavors). A flavor label did not include the brand names—for example, Brand X's "Cinnamon" and Brand Y's "Cinnamon" were treated as the same flavor. The main flavors coded were as follows: tobacco, menthol, alcohol/drinks, fruit, and dessert/candy. About 4.5% of flavors did not fall into these main categories. The flavors were generally coded by their first ingredient, with 2 exceptions: flavors that referenced tobacco were coded as tobacco and those described as minty, icy, or frosty were coded as menthol. Do-it-yourself flavor concentrates were excluded from this coding. We studied the proportion of the total flavors sold per brand in each of the main flavor categories.

Nicotine Strengths

For each website, research assistants recorded all distinct nicotine strengths listed. We reported the number of distinct strengths sold as well as whether each website offered zero-nicotine or nicotine-free options.

Statistical Analysis

Chi-square tests were conducted to determine significant differences in the types of product models sold, claims advanced by different brand types, and differences in flavor types sold. The McNemar test for paired data was used to examine changes in brands' product model, claims, and flavors sold from 2013-204 to 2016-2017. Moreover, *t* tests were used to examine differences in mean product model count and mean distinct flavor count. STATA version 14.2 (StataCorp LLC) was used for all analyses. Since 3 separate tests were performed (to compare tobacco-owned, internet-only, and vape shops to each other), we used a *P* value of .01 instead of the more conventional .05 for determining statistical significance.

Results

Brands and Models

A total of 178 (38.2%) of the 466 brands included in the 2013-2014 survey were no longer in operation as of July 2016, indicating substantial attrition in websites from 2013-2014 to 2016-2017. The 288 still operating in 2016-2017 provide a subset for longitudinal comparison. A total of 145 additional brands were uncovered through the 2016-2017 internet search. In total, we coded 433 websites in the 2016-2017 survey. We determined that 12 brands were owned by major tobacco companies, including *Blu* (Imperial Tobacco), *Logic* (Japan Tobacco), *MarkTen* (Altria Group, Inc.), and *VUSE* (Reynolds American, Inc.). Among the remaining brands, 162 operated their own physical vape shop and 259 were internet-only operations.

Table 1. A comparison of products and models offered by e-cigarette brands.

Type	2016-2017 survey (433 brands)						Evolution over time (288 brands)		
	Major tobacco (N=12), n (%)	Internet only (N=259), n (%)	Vape shop (N=162), n (%)	Tobacco vs internet, *P* value[a]	Tobacco vs vape shop, *P* value	Internet vs ape shop, *P* value	2013-2014 survey (N=288), n (%)	2016-2017 survey (N=288), n (%)	*P* value
Cigalike	10 (83.3)	143 (55.2)	47 (29.0)	.06	<.001	<.001	202 (70.1)	155 (53.8)	<.001
eGo	7 (58.3)	198 (76.4)	132 (81.5)	.15	.05	.22	208 (72.2)	227 (78.8)	.03
Mod	2 (16.7)	155 (59.8)	151 (93.2)	.003	<.001	<.001	117 (40.6)	190 (66.0)	<.001
Number of models, mean[b]	2.8	12.3	30.0	.08	<.001	<.001	6.7	17.2	<.001

[a]Comparisons between percentages were calculated using chi-square analysis in columns 1-6 and using McNemar test for paired data in columns 7-9.

[b]Comparisons between means were calculated with two-tailed *t* tests.

Table 1 shows data on product models sold online in 2016-2017. The first 3 columns list the percentage of brands in each type that sold cigalike, eGos, and mod designs through their websites. Columns 4-6 show *P* values for differences between brand types.

Cigalikes were sold by 83.3% (10/12) of the major tobacco company-owned brands compared with 55.2% of internet-only brands (143/259, *P*=.06) and 29.0% of vape shop brands (47/162, *P*<.001). Meanwhile, 16.7% (2/12) of tobacco-owned brands sold mods. This is significantly lower than the internet brands (59.8%, 155/259, *P*<.001) and vape shop brands (93.2%, 151/162, *P*<.001). Tobacco brands sold fewer product models on average (2.8) compared with internet (12.3, *P*=.08) and vape shop brands (30.0; *P*<.001). Overall, tobacco-owned brands and vape shop brands show the widest differences in products and models offered.

The last 3 columns of Table 1 compare the change in product types sold in 2013-2014 versus 2016-2017 among the 288 brands captured in both surveys. Brands became less likely to sell cigalikes (*P*<.001) and more likely to sell mods (*P*<.001) by 2016-2017. They also significantly increased their average number of product models sold from 6.7 to 17.2 (*P*<.001).

Claims and Disclaimers About Electronic Cigarettes

The columns 1-3 in Table 2 compare marketing claims made by different types of brand websites in 2016-2017. Overall, a low proportion of brands advanced direct claims regarding e-cigarettes as a method for smoking cessation, and none of the 12 tobacco-owned brands did so. Brands were more likely to advance indirect than direct claims. There were no significant differences between tobacco-owned brands versus vape shop and internet-only brands in terms of proclivity to advance smoking cessation claims, health-related claims, or social claims. Between internet and vape shop brands, the former were more likely to advance the social claims of e-cigarettes as being allowed in more places (*P*<.001) and being cleaner (*P*<.001) when compared with combustible cigarettes.

The majority of brands advanced disclaimers regarding e-cigarettes as not intended for youths/minors. Tobacco company-owned brands and internet-only brands show the greatest difference here. Moreover, 83.3% (10/12) of the tobacco-owned brands had an age-based pop-up window—a higher proportion than internet-only brands (50.2%, 130/259, *P*=.01). This proportion was also higher than the vape shop brands (60.5%, 98/162), but not statistically significant (*P*=.06).

The last 3 columns of Table 2 examine change in websites' claims from 2013-2014 to 2016-2017. Of the 6 claims coded in both surveys, 4 showed significant change. Although 57.3% (165/288) of brands advanced indirect smoking cessation claims in 2013-2014, less than half of this percentage (21.9%, 63/288) advanced them in 2016-2017 (*P*<.001). Brands also became significantly less likely to advance claims that e-cigarettes were healthier (*P*<.001), less expensive (*P*<.001), and could be smoked in more places (*P*<.001) than combustible cigarettes.

Flavors

The total count of distinct flavors sold by websites studied in 2016-2017 was 15,586—more than double the 7764 flavor labels found in 2013-2014. Major tobacco companies were less likely to sell e-liquids (66.7%, 8/12 sold e-liquids) relative to the internet-only (86.9%, 225/259, *P*=.05) and vape shop brands (98.1%, 159/162, *P*<.001). Table 3 (columns 1-3) compares flavors sold in 2016-2017 by brand type. Major tobacco companies sold fewer flavors on average (20.7 flavors) through their websites relative to vape shop brands (137.5 flavors, *P*=.002). Tobacco-owned companies sold significantly higher mean proportions of tobacco and menthol flavors relative to the internet and vape shop brands (all *P* values <.001). Conversely, tobacco-owned brands sold significantly lower proportions of alcohol/drink, fruit, and dessert/candy flavors relative to vape shop brands (all *P* values <.01).

Brands present in both survey periods became more likely to sell e-liquids (or prefilled cartridges)—89.6% (258/288) sold e-liquids (or prefilled cartridges) in 2016-2017 compared to 75.7% (218/288) in 2013-2014 (*P*<.001). The last 3 columns of Table 3 show other changes in the brands present in both survey periods. From 2013-2014 to 2016-2017, brands significantly decreased their proportions of tobacco (*P*<.001), menthol (*P*=.009), and alcohol/drink flavors (*P*=.003), and significantly increased their proportion of dessert/candy (*P*<.001) flavors. They also increased their average count of distinct flavors sold—from 49.2 to 81.6 (*P*<.001).

Table 2. A comparison of claims and disclaimers made by e-cigarette brands.

Claim	2016-2017 survey (433 brands)[a]						Evolution over time (288 brands)[b]		
	Major tobacco (N=12), n (%)	Internet only (N=259), n (%)	Vape shop (N=162), n (%)	Tobacco vs internet, *P* value	Tobacco vs vape shop, *P* value	Internet vs vape shop, *P* value	2013-2014 survey (N=288), n (%)	2016-2017 survey (N=288), n (%)	*P* value
Smoking cessation claims									
Direct: help quit	0 (0.0)	19 (7.3)	14 (8.6)	.33	.29	.63	29 (10.1)	32 (11.1)	.74
Indirect: help quit	2 (16.7)	50 (19.3)	23 (14.2)	.82	.81	.18	165 (57.3)	63 (21.9)	<.001
Not smoking cessation device	5 (41.7)	160 (61.8)	89 (54.9)	.16	.37	.17	160 (55.6)	177 (61.5)	.07
Health claims									
Healthier than smoking	5 (41.7)	129 (49.8)	64 (39.5)	.58	.88	.04	201 (69.8)	151 (52.4)	<.001
Social claims									
Less expensive than smoking	2 (16.7)	116 (44.8)	52 (32.1)	.06	.27	.01	178 (61.8)	125 (43.4)	<.001
Used in more places	6 (50.0)	119 (45.9)	42 (25.9)	.78	.07	<.001	174 (60.4)	132 (45.8)	<.001
Cleaner than smoking	3 (25.0)	132 (51.0)	54 (33.3)	.08	.55	<.001	NA[c]	NA	NA
Socially accepted	1 (8.3)	32 (12.4)	11 (6.8)	.68	.84	.07	NA	NA	NA
Age claims									
Disclaimer: not for minors	11 (91.7)	202 (78.0)	125 (77.2)	.26	.24	.84	NA	NA	NA
Age pop-up window	10 (83.3)	130 (50.2)	98 (60.5)	.01	.06	.02	NA	NA	NA

[a]Comparison in columns 1-6 were calculated using chi-square analysis.

[b]Comparisons in columns 7-9 were calculated using McNemar test for paired data.

[c]NA: Not available. Since variable was not coded for in the 2013-2014 survey, a longitudinal comparison is not possible.

Table 3. Flavors offered by e-cigarette brands.

Variable	2016-2017 survey (417 brands)						Evolution over time (278 brands)		
	Major tobacco (N=10)	Internet only (N=245)	Vape shop (N=162)	Tobacco vs internet, *P* value	Tobacco vs vape shop, *P* value	Internet vs vape shop, *P* value	2013-2014 survey (N=278)	2016-2017 survey (N=278)	*P* value
Number of flavors per brand, mean	20.7	56.3	137.5	.04	.002	<.001	49.2	81.6	<.001
Mean proportion of total flavors per brand, %[a]									
Tobacco	33.2	19.1	11.3	.005	<.001	<.001	21.9	18.1	<.001
Menthol	27.2	12.9	9.3	<.001	<.001	<.001	13.9	12.4	.009
Alcohol/Drink	7.6	12.5	12.6	.05	.001	.45	13.9	12.3	.003
Fruit	18.2	29.1	27.7	.01	<.001	.14	28.0	29.0	.13
Dessert/Candy	12.1	21.4	35.0	.03	<.001	<.001	16.7	23.9	<.001

[a]Comparisons between means and proportions were calculated with two-tailed *t* tests.

Table 4. Nicotine options offered by e-cigarette brands.

Number of nicotine options per brand	2016-2017 survey (407 brands)						Evolution over time (273 brands)		
	Major tobacco (N=10)	Internet only (N=237)	Vape shop (N=160)	Tobacco vs internet, *P* value	Tobacco vs vape shop, *P* value	Internet vs vape shop, *P* value	2013-2014 survey (N=273)	2016-2017 survey (N=273)	*P* value
Mean[a]	4.6	5.9	7.2	.09	.006	<.001	4.9	6.3	<.001
Offers zero nicotine[b], n (%)	5 (50.0)	211 (89.0)	156 (97.5)	<.001	<.001	.002	248 (90.8)	252 (92.3)	.57

[a]Comparisons between means were calculated with two-tailed *t* tests.

[b]Comparisons between percentages were calculated using chi-square analysis in columns 1-6 and using McNemar test for paired data in columns 7-9.

Nicotine Strengths

Table 4 shows that, in 2016-2017, tobacco company-owned brands offered fewer nicotine options on average relative to the vape shop brands (*P*=.006) and were less likely to offer zero-nicotine options compared with both vape shops and internet-only brands (*P* values <.001). Almost all the vape shop brands (97.5%, 156/160) and 89.0% (211/237) of internet brands offered a zero-nicotine option, whereas only half of the tobacco-owned brands (50.0%) offered one. The last 3 columns show that, from 2013-2014 to 2016-2017, brands significantly increased their average number of different nicotine options (*P*<.001). There was no change in the likelihood of a zero-nicotine option.

Discussion

Principal Findings

The analyses presented here show that major tobacco company-owned e-cigarette brands and smaller, independent e-cigarette brands offered very different portfolios of products through their online retail websites. Vape shop and tobacco company-owned brands appeared the most distinct from one another along a number of dimensions. Compared with vape shop brands, tobacco-owned brands were more likely to sell closed-system designs, were less likely to offer e-liquids, and tended to offer a narrower range of e-liquid flavors with greater emphasis on traditional (ie, tobacco and menthol) flavors. These are all product characteristics that resemble combustible cigarettes in appearance and taste. Vape shop brands, in comparison, were more likely to sell open-system models, focused more on nontraditional e-liquid flavors such as fruit and dessert/candy, offered a greater range of nicotine options, and were more likely to offer nicotine-free e-liquid. Internet-only brands, generally, resembled vape shop brands in their product offerings, although their differences with major tobacco brands were, generally, smaller in magnitude and less statistically significant with regard to product models, flavors, and nicotine options.

These patterns are consistent with research that suggests vape shops focus primarily on newer generation devices and encourage users' experimentation with a variety of nicotine and e-liquid options [20-22]. Studies also indicate vape shop employees frequently characterize e-cigarettes as smoking cessation devices and, despite lack of formal training, provide counsel regarding smoking cessation to customers [8,23]. Although we found a slightly higher rate of vape shops that claimed direct smoking cessation benefits to e-cigarettes relative to major tobacco brands, this difference was not statistically significant. Given the proliferation of vape shops in the United States, further study of this channel is important to understand the changing economic, social, and cultural dynamics of the e-cigarette market.

Among the 288 brand websites studied in both the 2013-2014 and 2016-2017, we found change in the propensity of e-cigarette brands to advance several types of claims about e-cigarettes. In 2013-2014, a majority of the websites presented indirect smoking cessation claims (57.3%, 165/288). This decreased to 21.9% (63/288) in 2016-2017. We also found significant decreases in claims regarding health and social benefits of e-cigarettes relative to combustible cigarettes. E-cigarette brands appear to have become more conservative in their marketing claims over time.

We also found, through our comparison of brands covered in both the 2013-2014 and 2016-2017 surveys, that brands have generally evolved from closed to open product designs, from traditional to nontraditional e-liquid flavors, and toward greater variety in models, flavors, and nicotine options. These general trends are important to keep in mind when considering the potential consequences of the 2016 Deeming Rule, which currently requires e-cigarette manufacturers to complete an application process that includes detailed ingredient, manufacturing, and product labeling/marketing information [24,25]. The FDA estimated that the resources to complete applications for new tobacco products will be considerable, costing between US $117,000 and US $466,000 per product (flavor-strength combination) [26].

A number of researchers and public health officials have raised concern that strict regulatory oversight will suppress product innovation; push out smaller, independent companies with limited resources; and ultimately strengthen market dominance for major tobacco companies [25,27]. Major tobacco companies, whose practices have been shaped by decades of experience with federal regulators, appear well-positioned and well-resourced to gain advantage within a strict regulatory environment. There is also some evidence suggesting that independent brands have already started reducing their product inventory and closing down operations in response to impending FDA oversight [28,29]. Our brand surveys indicate substantial

attrition in e-cigarette brands over the past few years. This contrasts sharply with the strong growth in e-cigarette brand websites found by Zhu et al from 2012 to 2014 [13], and it suggests that the industry may be headed toward even greater contraction in the number of brands in the coming years.

Some of the concerns regarding how regulatory oversight will affect the market landscape appear driven by tobacco companies' history of deceptive business practices and potential for heavily resourced, sophisticated marketing campaigns [11,30]. Our analyses suggest an additional reason for concern. Tobacco-owned brands are more likely to offer a limited range of product designs and flavors that closely resemble the experience of smoking combustible cigarettes. Although science regarding the harm reduction impact of e-cigarettes is still developing, some studies suggest that smokers of cigalikes and traditional tobacco/menthol flavors may be less likely to quit smoking and more likely to remain dual users for prolonged periods of time, which ultimately may be worse for their health outcomes [31,32]. There is also evidence that use of open-system e-cigarette models is associated with higher rates of smoking cessation relative to closed-system models [4,33,34]. This raises the possibility that greater dominance by major tobacco-owned brands and the exit of smaller, independent brands might ultimately limit smokers' access to e-cigarette models associated with higher quit rates.

On the other hand, we find major tobacco brands are more likely than internet brands to feature a pop-up window asking users to verify their age—83% (10/12) of tobacco brands compared with 50.2% (130/259) of internet and 60.5% (98/162) of vape shop brands. The latter 2 percentages are roughly comparable to recent content analyses of the internet e-cigarette vendors that studied age self-verification practices [18,19]. Of course, this represents a very weak form of age verification, and studies suggest the use of more effective verification methods, such as requiring a driving license number, which is relatively rare [19]. Still, these differences between major tobacco, internet-only, and vape shop brands suggest that major tobacco-owned brands may be more likely than small, independent brands to institute processes preventing online sales to minors. This raises the possibility that the exit of smaller brands could decrease youth access to e-cigarettes. Furthermore, tobacco-owned brands may also be less appealing to youths, as they offered limited ranges of flavors focused on tobacco and menthol, and studies suggest younger smokers find nontobacco flavors, such as fruit and dessert, appealing [35,36].

Limitations

One limitation of this study is that the search was limited to the first 30 pages of each keyword search. Other e-cigarette brands exist that are not in the database; thus, the total number of e-cigarette brands is likely larger than that reported here [37]. Another limitation is that we were only able to find limited information on tobacco industry ownership of e-cigarette brands. There may be other companies that we are unaware of that are tobacco-industry owned. Finally, as our study focuses on brands' websites, brands with physical retail stores could present different products and claims through their stores than those captured in this study.

Conclusions

The FDA has announced its intention to develop ENDS industry regulations that focus on nicotine and promote harm reduction through innovation [17,38]. As regulators consider how to best revise and implement oversight, there is a complex set of issues to take into account. Policies should be designed to discourage youths from starting e-cigarettes and exposing themselves to any potentially negative health effects of nicotine [39,40]. Currently, regulatory requirements do not provide clear guidance regarding specific processes needed to effectively prevent sales to minors [19]. In the 2016-2017 period, internet-only and vape shops brands appeared to take weaker actions to restrict youth access relative to major tobacco-owned brands. At the same time, enough preliminary evidence of a positive association between e-cigarette use and smoking cessation at the population level exists [31,41,42] that the FDA should consider how to encourage companies to continue to develop and offer a range of products. Regulations that involve intensive time- and resource-investments are likely to impose a disproportionate burden on small, independent organizations that have played a key role in developing and offering open-systems models that have been associated with higher quit rates [4,33,34]. Attention should be given to encouraging responsible business practices—particularly with regard to youth access regulation—while streamlining product application requirements to encourage product diversity and innovation among a variety of industry players.

Acknowledgments

The authors thank Zenjief del Castillo, Jeein Kim, and Madhur Singh for assistance with website coding.This work was supported by the National Cancer Institute of the National Institutes of Health under the State and Community Tobacco Control Initiative (grant number U01CA154280). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funder had no role in the decision to submit for publication or in review or approval of the manuscript for publication.

Conflicts of Interest

None declared.

References

1. Wells Fargo Securities. Nielsen: Tobacco “All Channel” Data Cig Pricing Remains Strong; E-Cig $ Sales Growth Re-Accelerates. In: Equity Research. San Francisco (CA): Wells Fargo Securities; Mar 31, 2015.
2. Farsalinos KE, Spyrou A, Tsimopoulou K, Stefopoulos C, Romagna G, Voudris V. Nicotine absorption from electronic cigarette use: comparison between first and new-generation devices. Sci Rep 2014 Feb 26;4:4133 [FREE Full text] [doi: 10.1038/srep04133] [Medline: 24569565]
3. Hajek P, Przulj D, Phillips A, Anderson R, McRobbie H. Nicotine delivery to users from cigarettes and from different types of e-cigarettes. Psychopharmacology (Berl) 2017 Mar;234(5):773-779 [FREE Full text] [doi: 10.1007/s00213-016-4512-6] [Medline: 28070620]
4. Chen C, Zhuang Y, Zhu S. E-cigarette design preference and smoking cessation: a U.S. population study. Am J Prev Med 2016 Sep;51(3):356-363. [doi: 10.1016/j.amepre.2016.02.002] [Medline: 27005984]
5. Brown CJ, Cheng JM. Electronic cigarettes: product characterisation and design considerations. Tob Control 2014 May;23(Suppl 2):ii4-i10 [FREE Full text] [doi: 10.1136/tobaccocontrol-2013-051476] [Medline: 24732162]
6. U.S. Department of Health and Human Services. Know the Risks. 2016. E-Cigarette use among youth and young adults: a report of the surgeon general URL: https://e-cigarettes.surgeongeneral.gov/documents/2016_sgr_full_report_non-508.pdf [accessed 2018-02-18] [WebCite Cache ID 6xL1ebd6l]
7. Lee YO, Kim AE. 'Vape shops' and 'E-Cigarette lounges' open across the USA to promote ENDS. Tob Control 2015 Jul;24(4):410-412. [doi: 10.1136/tobaccocontrol-2013-051437] [Medline: 24727649]
8. Sussman S, Baezconde-Garbanati L, Garcia R, Barker DC, Samet JM, Leventhal A, et al. Commentary: forces that drive the vape shop industry and implications for the health professions. Eval Health Prof 2016 Sep;39(3):379-388 [FREE Full text] [doi: 10.1177/0163278715586295] [Medline: 25967071]
9. Dai H, Hao J. Geographic density and proximity of vape shops to colleges in the USA. Tob Control 2017 Jul;26(4):379-385. [doi: 10.1136/tobaccocontrol-2016-052957] [Medline: 27302700]
10. Wells Fargo Securities. Tobacco talk: independent vapor Mfr survey: more bullish than retailers—suggesting vapor growth remains robust. In: Equity Research. San Francisco (CA): Wells Fargo Securities; Jan 23, 2015.
11. Grana RA, Ling PM. “Smoking revolution”: a content analysis of electronic cigarette retail websites. Am J Prev Med 2014 Apr;46(4):395-403 [FREE Full text] [doi: 10.1016/j.amepre.2013.12.010] [Medline: 24650842]
12. Klein EG, Berman M, Hemmerich N, Carlson C, Htut S, Slater M. Online e-cigarette marketing claims: a systematic content and legal analysis. Tob Regul Sci 2016 Jul;2(3):252-262 [FREE Full text] [doi: 10.18001/TRS.2.3.5] [Medline: 27446984]
13. Zhu SH, Sun JY, Bonnevie E, Cummins SE, Gamst A, Yin L, et al. Four hundred and sixty brands of e-cigarettes and counting: implications for product regulation. Tob Control 2014 Jul;23(Suppl 3):iii3-iii9 [FREE Full text] [doi: 10.1136/tobaccocontrol-2014-051670] [Medline: 24935895]
14. Seidenberg AB, Jo CL, Ribisl KM. Differences in the design and sale of e-cigarettes by cigarette manufacturers and non-cigarette manufacturers in the USA. Tob Control 2016 Apr;25(e1):e3-e5. [doi: 10.1136/tobaccocontrol-2015-052375] [Medline: 26546151]
15. FDA. Vaporizers, e-Cigarettes, and other electronic nicotine delivery systems (ENDS) URL: https://www.fda.gov/TobaccoProducts/Labeling/ProductsIngredientsComponents/ucm456610.htm#regulation [accessed 2017-11-19] [WebCite Cache ID 6sP6rcxuR]
16. Abbasi J. FDA extends authority to e-Cigarettes: implications for smoking cessation? J Am Med Assoc 2016 Aug 09;316(6):572-574. [doi: 10.1001/jama.2016.8568] [Medline: 27419929]
17. FDA. FDA announces comprehensive regulatory plan to shift trajectory of tobacco-related disease, death URL: https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/UCM568923.htm [accessed 2017-11-19] [WebCite Cache ID 6stGuo2dA]
18. Mackey TK, Miner A, Cuomo RE. Exploring the e-cigarette e-commerce marketplace: identifying internet e-cigarette marketing characteristics and regulatory gaps. Drug Alcohol Depend 2015 Nov 01;156:97-103. [doi: 10.1016/j.drugalcdep.2015.08.032] [Medline: 26431794]
19. Williams RS, Derrick J, Liebman AK, LaFleur K, Ribisl KM. Content analysis of age verification, purchase and delivery methods of internet e-cigarette vendors, 2013 and 2014. Tob Control 2017 May 08:- Epub ahead of print(forthcoming). [doi: 10.1136/tobaccocontrol-2016-053616] [Medline: 28484040]
20. Burbank AD, Thrul J, Ling PM. A pilot study of retail 'Vape Shops' in the San Francisco bay area. Tob Prev Cessat 2016;2(Suppl):6 [FREE Full text] [doi: 10.18332/tpc/65229] [Medline: 28393129]
21. Allem JP, Unger JB, Garcia R, Baezconde-Garbanati L, Sussman S. Tobacco attitudes and behaviors of vape shop retailers in Los Angeles. Am J Health Behav 2015 Nov;39(6):794-798 [FREE Full text] [doi: 10.5993/AJHB.39.6.7] [Medline: 26450547]
22. Tsai JY, Bluthenthal R, Allem JP, Garcia R, Garcia J, Unger J, et al. Vape shop retailers' perceptions of their customers, products and services: a content analysis. Tob Prev Cessat 2016;2(Suppl):3 [FREE Full text] [doi: 10.18332/tpc/70345] [Medline: 28736758]
23. Cheney MK, Gowin M, Wann TF. Vapor store owner beliefs about electronic cigarette regulation. Tob Regul Sci 2015 Oct 01;1(3):227-235. [doi: 10.18001/TRS.1.3.4]

24. FDA. 2017. Effective and compliance dates applicable to retailers, manufacturers, importers, and distributors of newly deemed tobacco products URL: https://www.fda.gov/downloads/TobaccoProducts/GuidanceComplianceRegulatoryInformation/UCM501016.pdf [accessed 2017-09-18] [WebCite Cache ID 6sP72RRkD]
25. Tavernise S. Nytimes. 2016 May 05. F.D.A. imposes rules for e-Cigarettes in a landmark move URL: https://www.nytimes.com/2016/05/06/science/fda-rules-electronic-cigarettes.html [accessed 2017-11-20] [WebCite Cache ID 6sP7FwSgq]
26. FDA. Questions on FDA's new regulations for E-cigarettes, cigars, and all other tobacco products URL: https://www.fda.gov/TobaccoProducts/AboutCTP/ucm378205.htm [accessed 2017-08-02] [WebCite Cache ID 6sP7T48ca]
27. Abrams DB. Promise and peril of e-cigarettes: can disruptive technology make cigarettes obsolete? J Am Med Assoc 2014 Jan 08;311(2):135-136. [doi: 10.1001/jama.2013.285347] [Medline: 24399548]
28. Ericksen AB. Cstoredecisions. 2017 Feb 15. E-tobacco: what's next? URL: https://www.cstoredecisions.com/2017/02/15/e-tobacco-whats-next/ [accessed 2017-09-13] [WebCite Cache ID 6sP7ipTFP]
29. WFYI. 2016. Indiana vape shop owners say new FDA rules will shut them down internet URL: http://www.wfyi.org/news/articles/indiana-vape-shop-owners-say-new-fda-rules-will-shut-them-down [accessed 2017-08-02] [WebCite Cache ID 6sP7s7Xrw]
30. Hsu G, Grodal S. Category taken-for-grantedness as a strategic opportunity: the case of light cigarettes. Am Sociol Rev 2015 Jan 29;80(1):28-62. [doi: 10.1177/0003122414565391]
31. Zhuang YL, Cummins SE, Sun JY, Zhu SH. Long-term e-cigarette use and smoking cessation: a longitudinal study with US population. Tob Control 2016 Dec;25(Suppl 1):i90-i95 [FREE Full text] [doi: 10.1136/tobaccocontrol-2016-053096] [Medline: 27697953]
32. Tackett AP, Lechner WV, Meier E, Grant DM, Driskill LM, Tahirkheli NN, et al. Biochemically verified smoking cessation and vaping beliefs among vape store customers. Addiction 2015 May;110(5):868-874. [doi: 10.1111/add.12878] [Medline: 25675943]
33. Hitchman SC, Brose LS, Brown J, Robson D, McNeill A. Associations between e-cigarette type, frequency of use, and quitting smoking: findings from a longitudinal online panel survey in Great Britain. Nicotine Tob Res 2015 Oct;17(10):1187-1194 [FREE Full text] [doi: 10.1093/ntr/ntv078] [Medline: 25896067]
34. Malas M, van der Tempel J, Schwartz R, Minichiello A, Lightfoot C, Noormohamed A, et al. Electronic cigarettes for smoking cessation: a systematic review. Nicotine Tob Res 2016 Dec;18(10):1926-1936. [doi: 10.1093/ntr/ntw119] [Medline: 27113014]
35. Audrain-McGovern J, Strasser AA, Wileyto EP. The impact of flavoring on the rewarding and reinforcing value of e-cigarettes with nicotine among young adult smokers. Drug Alcohol Depend 2016 Dec 01;166:263-267 [FREE Full text] [doi: 10.1016/j.drugalcdep.2016.06.030] [Medline: 27426010]
36. Goldenson NI, Kirkpatrick MG, Barrington-Trimis JL, Pang RD, McBeth JF, Pentz MA, et al. Effects of sweet flavorings and nicotine on the appeal and sensory properties of e-cigarettes among young adult vapers: application of a novel methodology. Drug Alcohol Depend 2016 Dec 01;168:176-180 [FREE Full text] [doi: 10.1016/j.drugalcdep.2016.09.014] [Medline: 27676583]
37. Williams RS, Derrick J, Ribisl KM. Electronic cigarette sales to minors via the internet. JAMA Pediatr 2015 Mar;169(3):e1563 [FREE Full text] [doi: 10.1001/jamapediatrics.2015.63] [Medline: 25730697]
38. Gottlieb S, Zeller M. A nicotine-focused framework for public health. N Engl J Med 2017 Sep 21;377(12):1111-1114. [doi: 10.1056/NEJMp1707409] [Medline: 28813211]
39. Pisinger C, Døssing M. A systematic review of health effects of electronic cigarettes. Prev Med 2014 Dec;69:248-260 [FREE Full text] [doi: 10.1016/j.ypmed.2014.10.009] [Medline: 25456810]
40. England LJ, Bunnell RE, Pechacek TF, Tong VT, McAfee TA. Nicotine and the developing human: a neglected element in the electronic cigarette debate. Am J Prev Med 2015 Aug;49(2):286-293 [FREE Full text] [doi: 10.1016/j.amepre.2015.01.015] [Medline: 25794473]
41. Beard E, West R, Michie S, Brown J. Association between electronic cigarette use and changes in quit attempts, success of quit attempts, use of smoking cessation pharmacotherapy, and use of stop smoking services in England: time series analysis of population trends. Br Med J 2016 Sep 13;354:i4645 [FREE Full text] [Medline: 27624188]
42. Zhu SH, Zhuang YL, Wong S, Cummins SE, Tedeschi GJ. E-cigarette use and associated changes in population smoking cessation: evidence from US current population surveys. Br Med J 2017 Jul 26;358:j3262 [FREE Full text] [Medline: 28747333]

Abbreviations

e-cigarette: electronic cigarette
ENDS: electronic nicotine delivery system
FDA: Food and Drug Administration

Edited by G Eysenbach; submitted 29.08.17; peer-reviewed by JP Allem, JC Chen, L Gundel; comments to author 14.11.17; accepted 06.12.17; published 12.03.18.

Please cite as:
Hsu G, Sun JY, Zhu SH
Evolution of Electronic Cigarette Brands From 2013-2014 to 2016-2017: Analysis of Brand Websites
J Med Internet Res 2018;20(3):e80
URL: http://www.jmir.org/2018/3/e80/
doi:10.2196/jmir.8550
PMID:29530840

Original Paper

Scope, Breadth, and Differences in Online Physician Ratings Related to Geography, Specialty, and Year: Observational Retrospective Study

Jessica Janine Liu[1*], MSc, MD; John Justin Matelski[2*], MSc; Chaim M Bell[3*], MD, PhD

[1]Department of Medicine, University of Toronto, University Health Network, Toronto, ON, Canada

[2]Biostatistics Research Unit, University Health Network, Toronto, ON, Canada

[3]Sinai Health System, Department of Medicine, University of Toronto, Toronto, ON, Canada

[*]all authors contributed equally

Corresponding Author:
Jessica Janine Liu, MSc, MD
Department of Medicine
University of Toronto
University Health Network
Toronto General Hospital, 14EN213
200 Elizabeth Street
Toronto, ON, M5G 2C4
Canada
Phone: 1 416 340 3111 ext 4908
Fax: 1 416 595 5826
Email: jessica.liu@uhn.ca

Abstract

Background: Physician ratings websites have emerged as a novel forum for consumers to comment on their health care experiences. Little is known about such ratings in Canada.

Objective: We investigated the scope and trends for specialty, geographic region, and time for online physician ratings in Canada using a national data source from the country's leading physician-rating website.

Methods: This observational retrospective study used online ratings data from Canadian physicians (January 2005-September 2013; N=640,603). For specialty, province, and year of rating, we assessed whether physicians were likely to be rated favorably by using the proportion of ratings greater than the overall median rating.

Results: In total, 57,412 unique physicians had 640,603 individual ratings. Overall, ratings were positive (mean 3.9, SD 1.3). On average, each physician had 11.2 (SD 10.1) ratings. By comparing specialties with Canadian Institute of Health Information physician population numbers over our study period, we inferred that certain specialties (obstetrics and gynecology, family practice, surgery, and dermatology) were more commonly rated, whereas others (pathology, radiology, genetics, and anesthesia) were less represented. Ratings varied by specialty; cardiac surgery, nephrology, genetics, and radiology were more likely to be rated in the top 50th percentile, whereas addiction medicine, dermatology, neurology, and psychiatry were more often rated in the lower 50th percentile of ratings. Regarding geographic practice location, ratings were more likely to be favorable for physicians practicing in eastern provinces compared with western and central Canada. Regarding year, the absolute number of ratings peaked in 2007 before stabilizing and decreasing by 2013. Moreover, ratings were most likely to be positive in 2007 and again in 2013.

Conclusions: Physician-rating websites are a relatively novel source of provider-level patient satisfaction and are a valuable source of the patient experience. It is important to understand the breadth and scope of such ratings, particularly regarding specialty, geographic practice location, and changes over time.

(J Med Internet Res 2018;20(3):e76) doi:10.2196/jmir.7475

KEYWORDS

quality improvement; patient satisfaction; patient-centered care; online ratings

Introduction

Patients' abilities to discern health care quality are often underappreciated, despite evidence that low patient satisfaction scores and complaints against physicians are linked to increased risk management episodes, malpractice lawsuits, readmission rates, and even increased mortality for selected diagnoses [1-5]. Over the last decade, physician-rating websites have become a popular source of patient satisfaction data [6]. Such websites represent unsolicited reflections of the patient experience with their physicians in comparison to more traditional methods such as surveys. In the United States, the use of physician-rating websites is rapidly increasing, whereas other countries have reported more moderate growth [6,7]. In addition to private online physician websites, government or health insurer-developed sites are also being used in countries such as the United Kingdom and Germany [8,9]. Together, these physician-rating websites may impact patient health care decision making, as data suggests approximately one-third of users have searched for physicians online and report making decisions regarding physician selection based on these ratings [10]. Online physician-rating websites may also impact physician behaviors; over the last five years, physicians have been increasingly responding online to their ratings [11]. Hence, this data source may have significant implications on health care practice and behavior.

Most previous work on online physician ratings has focused on reviewing the frequency and usage among different physician specialties in the United States, China, and Germany [6,12-26], as well as exploring awareness and perceptions among physicians and consumers [10,11,27-29]. More recently, the focus has been to correlate online ratings with quality outcomes or surrogates such as postoperative mortality and surgical volumes with variable findings, depending on the quality outcome in question [6,30-38]. It has been estimated that one in six physicians are rated online, and most ratings are positive [6,12-4,17-19,28]. Although the use of physician-rating websites is increasing overall [6,7,39], for frequency of ratings, US studies have reported that the mean number of ratings per physician is low overall, ranging from two to four ratings per physician [6,17,21]. Several studies have focused on differences in ratings according to specialties [6,14,20-22]. Certain types of physicians, such as obstetricians, dermatologists, surgeons, and family physicians, are more frequently rated than other specialists. Board-certified, younger physicians have been shown to be rated more favorably than non-board-certified, younger physicians [6,14]. Other studies have investigated the relationship between practice location (such as city size) and online ratings. In the United States, physicians in the southern states had a higher likelihood of positive ratings than other parts of the country, whereas others have shown no difference in ratings with respect to practice location and city size [6,20-22].

In Canada, there is currently little information available on the use of physician-rating websites. Our study sought to investigate the nature and trends of online physician ratings in Canada over a nearly 8-year period. The goals of this study were to (1) determine whether online ratings for physicians differed depending on physician specialty, (2) investigate whether physician practice location affected online ratings, and (3) examine possible trends in ratings over time by year of rating. We also compared the number and frequency of ratings by specialty to determine whether certain specialties were rated online more frequently than expected based on their representation in the overall physician population. Based on previous studies, we hypothesized that certain specialties, such as obstetrics and family medicine, would be rated more frequently than others, such as pathology or radiology. We also felt that the quality of ratings would be positive overall and that differences in ratings would exist across specialties and geographic practice location. We suspected that there would be no differences in quality of ratings over time, but that the absolute number of ratings would be steadily increasing over our study period.

Methods

Overview

We accessed a national database of all Canadian physicians rated from January 2005 to September 2013 (N=640,603 ratings) [40]. RateMDs was founded in the United States in 2004 and is currently among the most popular physician-rating websites in Canada and the United States by user traffic [13,18]. No registration or subscription is required to view or post a rating, and there are no incentives to rate a physician. Physicians are rated on a scale of 1 to 5 (described by the website as 1="terrible," 2="poor," 3="okay," 4="good," 5="excellent"). Ratings were given for each of the following domains: staff, punctuality, helpfulness, and knowledge. A mean overall score is posted for each physician. Physician profiles are created or searched for by the rater, and users provide ratings and may provide free-text comments if desired. Our dataset included deidentified data for 57,412 physicians, including specialty, practice region (city and province), date of rating, and scores on each of four domains, from which we calculated an average cumulative rating for each physician. This dataset included all physicians in Canada who were rated on RateMDs during our study period.

Mean number of ratings and mean ratings were calculated for all physicians, each website specialty, and province. To compare the relative proportions of physicians by specialty, we grouped specialties according to Canadian Institute of Health Information (CIHI) categories [41]. We considered "obstetrics and gynecology" as distinct from "surgery" because previous research demonstrated high numbers of ratings for this group [6,14]. We calculated each physician specialty's online presence by grouping online specialties into CIHI specialty categorizations and divided the number of physicians rated online for that specialty by the total number of physicians in the online database. We then calculated and compared these values to the mean annual number of physicians divided by the total annual physician population for CIHI specialties from 2005 to 2013 (to match our online ratings data period). This allowed us to infer whether a specialty was rated more or less frequently than expected based on the mean annual physician population for that specialty.

Statistics

For statistical analysis, our objective was to recognize and compare differences in favorable versus unfavorable ratings for physician specialty, geographic practice location, and year of rating. We constructed a binary variable indicating whether each rating was greater than, less than, or equal to the median rating, which was 4.5 out of 5. We thus considered whether ratings were in the top 50th percentile of all ratings for one of three predictors: physician specialty, province, and year. For each level of predictor, the proportion of ratings greater than 4.5 was reported with a 95% confidence interval and a *P* value against the null hypothesis that the true proportion was equal to 0.5. In this way, we were able to stratify specialties, practice location by province, and year of rating according to likelihood of positive ratings. All analyses were performed using *prop.test* in R version 3.0.2.

Ethics

When submitting research ethics board approval, we were informed that the requirement for ethics approval was waived because data were publicly available.

Results

Findings

From February 2005 to September 2013, there were 640,603 ratings for 57,412 unique physicians. Ratings were generally positive (mean 3.9, SD 1.3). Using the online rating website's rating descriptions, this translated to a mean rating that fell between "okay" and "good." The mean number of ratings per physician was 11.2 (SD 10.1; see Table 1). During our study period, the mean annual number of total physicians in Canada was 66,026.1 (SD 5748.2). The largest group of physicians, by medical specialty, was family medicine/general practice (n=30,818 physicians). This group had 370,972 unique ratings and, on average, had 12 ratings per physician, with a mean overall rating of 3.9 (SD 1.3). Internal medicine (including its subspecialties) accounted for 53,818 total ratings of 6677 individual physicians, with 8.1 ratings per physician (SD 7.7) and a mean rating of 3.98 (SD 1.31) out of 5. Surgery (including its subspecialties) included 22,811 total ratings of 2472 individual physicians, with 11.9 ratings per physician (SD 10.7) and an overall mean rating of 4.01 (SD 1.32) out of 5. We found that certain specialties had relatively increased numbers of per-physician ratings, including reproductive endocrinology (mean 19.7, SD 15.2), cosmetics/plastic surgery (mean 16.7, SD 16.1), and obstetrics and gynecology (mean 17.6, SD 16.1). Additionally, certain medical specialties had lower numbers of rated physicians as well as per-physician ratings, including radiologists (total number of rated physicians: 330, mean per-physician ratings 3.0, SD 2.8), pathologists (total number of rated physicians: 13, mean per-physician ratings 4.4, SD 8.1), and medical geneticists (total number of rated physicians: 26, mean per-physician ratings 2.6, SD 4.9; see Table 1).

Differences in Frequencies of Ratings According to Specialty

For each specialty, we calculated the percentage of physicians with online ratings divided by the total online physician population, and compared it to the percentage of physicians in a given specialty divided by the total annual physician populations for CIHI specialties. Certain specialties were more frequently rated than expected based on their proportion in the national population, notably obstetrics and gynecology (4.3% of online cohort vs 2.5% of mean total annual obstetrics and gynecology population), dermatology (1.2% vs 0.8%), family practice (53.7% vs 45.2%), internal medicine (including its subspecialties; 12.0% vs 10.3%), emergency medicine/critical care (2.4% vs 1.1%), and surgery (14.3% vs 10.0%), whereas others were less represented, including anesthesia (1.4% vs 4.1%), radiology (0.6% vs 3.3%), psychiatry (5.0% vs 6.3%), and pathology (<0.01% vs 1.9%; see Table 1).

Differences in Quality of Ratings for Physician Specialty

We investigated whether there were differences in the quality of ratings depending on physician specialty. We found that ratings for certain specialties were more likely to be in the top 50th percentile of all ratings, including cardiac surgery (probability of a rating greater than the median of 4.5 was 78.1%, $P<.001$), genetics (73.5%, $P<.001$), nephrology (69.2%, $P<.001$), radiology (65.3%, $P<.001$), and vascular surgery (65.1%, $P<.001$). The bottom four physician specialties included psychiatry (42.2%, $P<.001$), neurology (42.1%, $P<.001$), dermatology (37.0%, $P<.001$), and addiction medicine (35.8%, $P<.001$; Figure 1; Multimedia Appendix 1). Family medicine/general practice comprised our largest group of physicians in the online cohort, as well as one of the largest groups of physicians represented in the mean annual physician population. Regarding likelihood of a favorable rating, family medicine/general practice was among the bottom seven physician specialties (46.3%, $P<.001$; Figure 1; Multimedia Appendix 1).

Differences in Frequency of Ratings for Physician Practice Location (by Province)

We found that Ontario had both the highest number of ratings and the highest number of rated unique physicians (244,635 ratings for 20,740 physicians), followed by Quebec (116, 041 for 13,460 physicians), then British Columbia (101,152 ratings for 8398 physicians). The lowest number of ratings for the lowest number of physicians was found in the less densely populated regions of the Northwest Territories/Yukon/Nunavut (802 ratings for 126 physicians) and Prince Edward Island (2534 ratings for 242 physicians).

For most provinces, per-physician number of ratings ranged from 10 to 13, with the exception of Quebec and the Northwest Territories/Yukon/Nunavut (ratings per physician 8.62 and 6.37, respectively).

Differences in Quality of Ratings for Physician Practice Location (by Province)

We also found differences in a physician's likelihood of a positive rating depending on practice location. Broadly speaking, physicians who practiced medicine in the eastern geographic locations of the country had a higher likelihood of being favorably rated than those who practiced in central or western Canada.

Table 1. Number of ratings, unique number of physicians, and descriptive statistics and relative proportions of rated physicians grouped by Canada Institute of Health Information (CIHI) specialty (2005-2013).

Medical specialty	Ratings, n	Unique rated physicians, n	Ratings per physician, mean (SD)	Overall rating, mean (SD)	Annual physician population[a] (%), mean (SD)	% of online cohort	% of mean annual physician population[b]
Family medicine	370,972	30,818	12.0 (9.7)	3.9 (1.3)	33,180.0 (4745.8)	53.7	45.2
Internal medicine	53,818	6677	8.1 (7.7)	4.0 (1.3)	7528.1 (778.3)	12.0	10.3
Allergy/immunologist	2690	235	11.5 (9.6)	3.8 (1.3)			
Cardiologist	8192	1278	6.4 (5.8)	4.20 (1.2)			
Colorectal/proctologist	312	36	8.7 (9.5)	4.02 (1.4)			
Gastroenterologist	9395	854	11.0 (8.8)	3.91 (1.3)			
Endocrinologist	5670	563	10.1 (8.6)	3.81 (1.3)			
Reproductive endocrinologist	1418	72	19.7 (15.2)	3.91 (1.3)			
Geriatrician	678	168	4.0 (4.4)	3.84 (1.4)			
Infectious disease	1074	199	5.4 (6.1)	4.04 (1.3)			
Internist	7045	1112	6.3 (6.4)	3.90 (1.4)			
Nephrologist	1868	364	5.1 (4.4)	4.36 (1.1)			
Oncology/hematologist	7038	1086	6.5 (6.1)	4.12 (1.2)			
Pulmonologist	2463	383	6.4 (5.8)	4.10 (1.3)			
Rheumatologist	5915	472	12.5 (9.4)	3.81 (1.4)			
Sleep disorders	371	57	6.5 (6.9)	3.7 (1.4)			
Anesthesia	2589	622	4.16 (5.0)	4.14 (1.3)	2713.1 (287.0)	1.1	4.1
Obstetrics and gynecology	43,627	2472	17.6 (15.0)	3.88 (1.3)	1709.0 (292.3)	4.3	2.5
Surgery	98,045	8235	11.9 (10.7)	4.01 (1.3)	6618.2 (375.3)	14.3	10.0
Surgeon (general)	22,811	2185	10.4 (9.1)	4.16 (1.3)			
Cardiothoracic surgeon	1954	202	9.6 (7.8)	4.54 (1.0)			
Cosmetic/plastics	13,226	793	16.7 (16.1)	4.05 (1.3)			
Otolaryngology	10,064	736	13.6 (11.4)	3.87 (1.4)			
Neurosurgeon	4686	363	12.9 (10.7)	4.17 (1.3)			
Ophthalmologist	12,419	1305	9.5 (8.5)	3.90 (1.3)			
Orthopedics/sport	22,492	1770	12.7 (10.0)	3.91 (1.4)			
Bariatric/weight loss	346	26	13.3 (30.5)	4.05 (1.4)			
Urologist	9655	772	12.5 (9.5)	4.00 (1.3)			
Vascular surgeon	392	83	4.7 (5.0)	4.13 (1.4)			
Neurology	9504	944	10.1 (9.5)	3.59 (1.4)	829.2 (83.1)	1.6	1.2
Pediatrics	20,751	1767	11.7 (10.7)	4.1 (1.2)	2488.0 (508.9)	3.1	3.8
Radiology	1005	330	3.0 (2.8)	4.18 (1.3)	2153.7 (216.2)	0.6	3.3
Emergency/critical care[c]	7716	1404	5.5 (5.4)	3.81 (1.5)	860.2 (260.0)	2.4	1.1
Psychiatry	18,036	2853	6.3 (6.4)	3.55 (1.5)	4218.9 (550.5)	5.0	6.3
Psychiatry (general)	17,695	2784	6.4 (6.4)	3.55 (1.5)			
Addiction medicine	341	69	4.9 (5.6)	3.49 (1.5)			
Dermatology	11,587	705	16.4 (14.9)	3.53 (1.4)	540.0 (27.0)	1.2	0.8
Pathology	57	13	4.4 (8.1)	4.18 (1.2)	1271.4 (100.1)	<0.01	1.9
Genetics	68	26	2.6 (4.9)	4.61 (0.8)	75.8 (12.6)	<0.01	<0.01

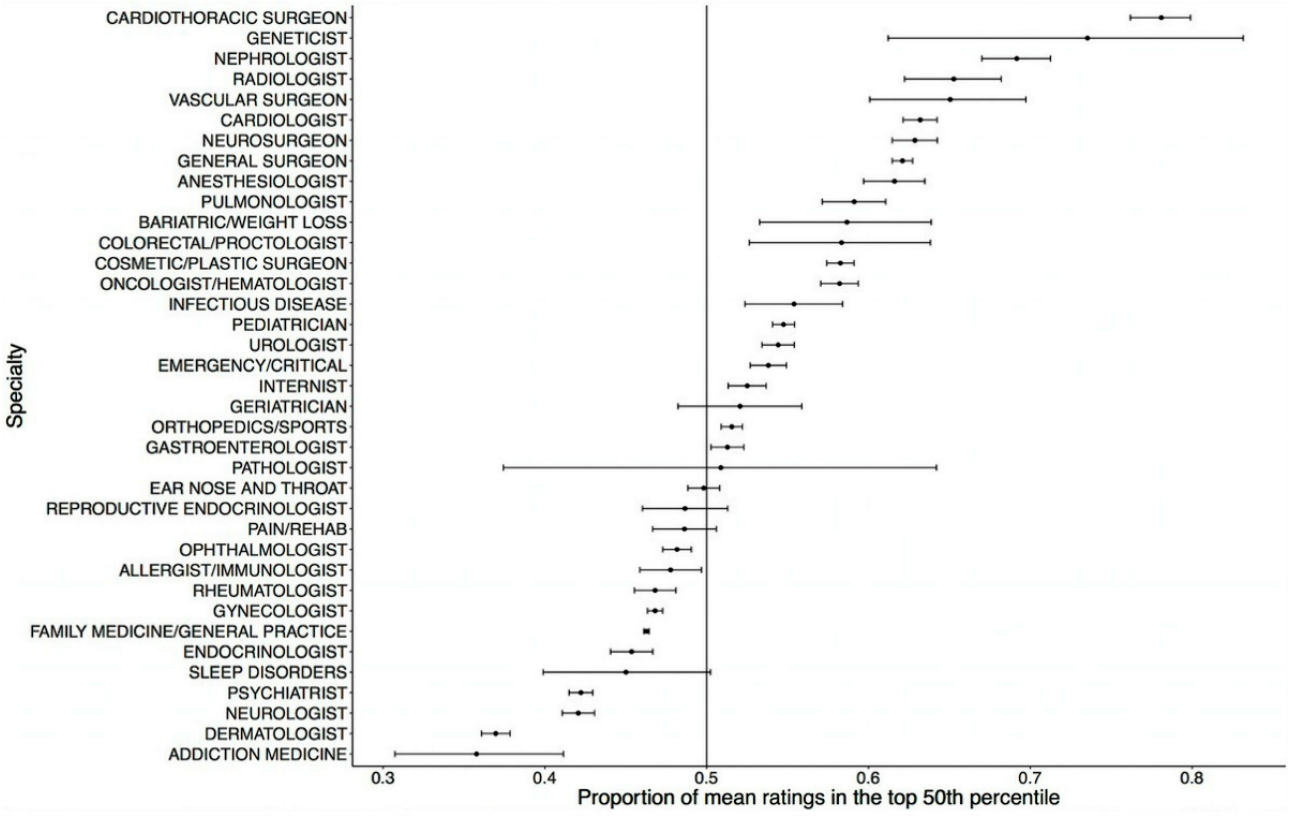
Medical specialty	Ratings, n	Unique rated physicians, n	Ratings per physician, mean (SD)	Overall rating, mean (SD)	Annual physician population[a] (%), mean (SD)	% of online cohort	% of mean annual physician population[b]
Physical medicine/rehabilitation	2517	344	7.3 (8.3)	3.71 (1.5)	372.3 (41.2)	0.6	0.6
Totals/means (SD)	640,603	57,412	11.2 (10.1)	3.9 (1.3)	66,026.1 (5748.2)		

[a]For each specialty, number of unique physicians rated online per total number of unique physicians rated online, expressed as a percent.

[b]For each CIHI physician specialty, mean annual number of physicians per mean total number of annual physicians (2005-2013) expressed as a percent.

[c]Emergency/critical care, as a grouped CIHI specialty, was only available for the years 2009-2013; therefore, annual means were calculated over 5 years only for this specialty.

Figure 1. Proportion of mean ratings, by specialty, in the top 50th percentile of all rated physicians (2005-2013) with 95% confidence intervals depicted for each proportion.



Specifically, physicians practicing in New Brunswick (56.3%, *P*<.001), Newfoundland (56.0%, *P*<.001), Quebec (53.6%, *P*<.001), Prince Edward Island (53.6%, *P*<.001), the Northwest Territories/Yukon/Nunavut (52.7%, *P*=.13), and Nova Scotia (52.7%, *P*<.001) were more likely to be rated greater than 4.5, whereas those practicing in Saskatchewan (46.4%, *P*<.001), Ontario (46.9%, *P*<.001), British Columbia (46.5%, *P*<.001), Alberta (46.5%, *P*<.001), and Manitoba (45.6%, *P*<.001) were likely to be rated 4.5 or lower (Figure 2; Multimedia Appendix 1).

Differences in Online Ratings for Year of Rating

During our study period, there were 640,603 total individual ratings of 27,181 physicians. Over time, the total number of ratings continued to increase; however, we found some important differences in the number of additional new ratings per year (Table 2). In 2005, when the website was still new in Canada, there were only 138 ratings. However, in 2007, 200,650 new ratings were posted before slowly tapering down each subsequent year until 2013, when there were 51,800 new ratings. The year 2007 was also notable in that the mean number of ratings per physician was highest at 5.74 (SD 5.28) before settling at 1 to 3 ratings per physician. In terms of quality of ratings, from 2005 to 2013, physicians were more likely to be rated above the median if rated more recently (ie, in 2013; upper 50th percentile proportion 0.512, *P*<.001), and the likelihood of favorable ratings increased over time. There were two years (2013 and 2007) when quality of ratings were especially high, whereas for the remaining years the proportion of ratings greater than 4.5 was significantly less than 50% (Figure 3; Multimedia Appendix 1).

Figure 2. Proportion of mean ratings, by province, in the top 50th percentile of all rated physicians (2005-2013) with 95% confidence intervals depicted for each proportion. NB: New Brunswick; NL: Newfoundland and Labrador; QC: Quebec: PE: Prince Edward Island; NT/YT/NU: Northwest Territories/Yukon/Nunavut; NS: Nova Scotia: SK: Saskatchewan; ON: Ontario; BC: British Columbia; AB: Alberta: MB: Manitoba.

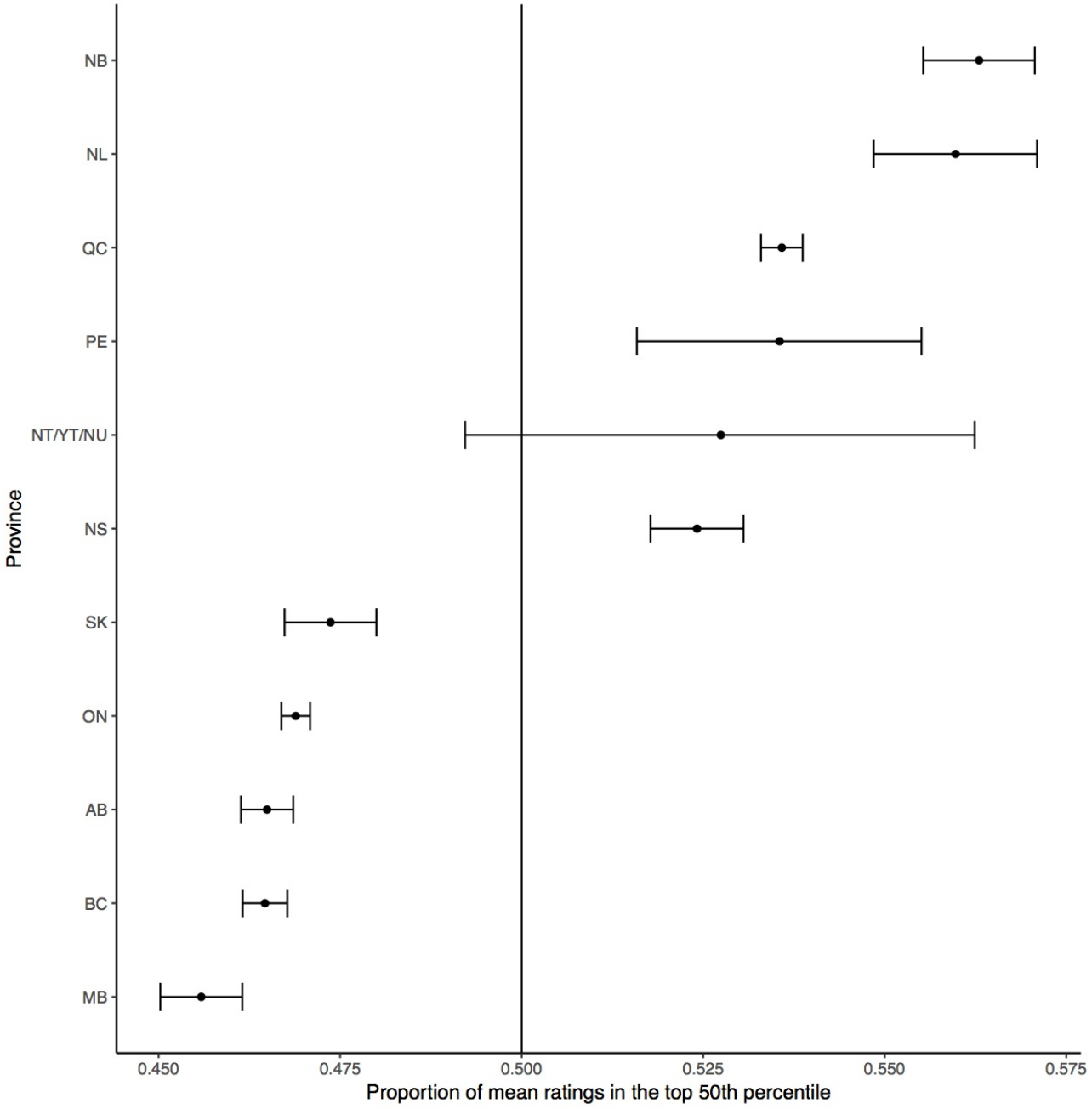
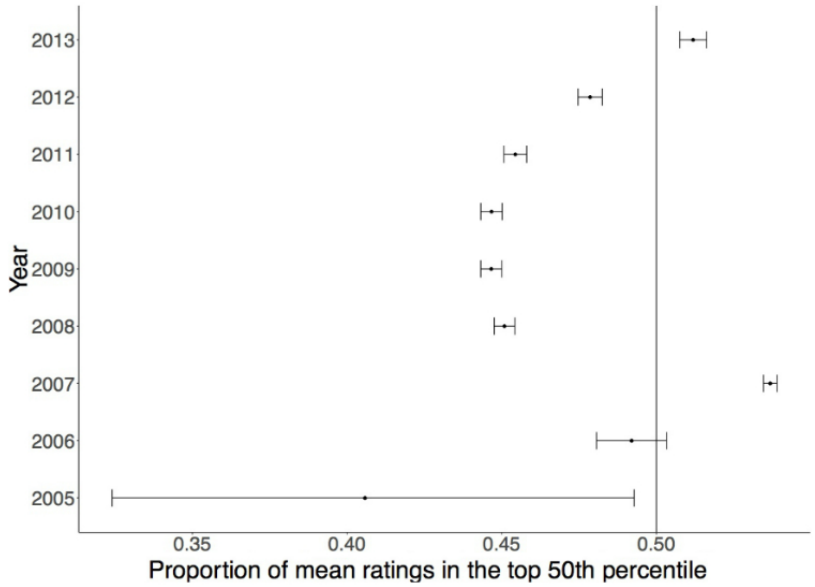


Table 2. Number of ratings, number of physicians, mean ratings per physician, mean overall rating, and additional ratings per year of all physicians rated on RateMDs by province and by year of rating (2005-2013).

Category	Ratings, n	Physicians, n	Ratings per physician, mean (SD)	Mean overall rating, mean (SD)	Additional ratings per year
Province					
New Brunswick	16,128	1447	11.15 (8.9)	4.03 (1.29)	—
Newfoundland	7564	893	8.47 (7.2)	4.09 (1.23)	—
Prince Edward Island	2534	242	10.47 (8.0)	4.00 (1.29)	—
Quebec	116,041	13,460	8.62 (8.5)	4.04 (1.28)	—
Northwest Territories/Yukon/Nunavut	802	126	6.37 (6.0)	3.94 (1.34)	—
Nova Scotia	23,482	1992	11.79 (9.5)	3.99 (1.28)	—
Saskatchewan	24,093	1880	12.82 (11.7)	3.84 (1.34)	—
Ontario	244,635	20,740	11.80 (10.4)	3.86 (1.33)	—
Alberta	74,077	5968	12.41 (11.1)	3.86 (1.32)	—
British Columbia	101,152	8398	12.04 (9.9)	3.87 (1.31)	—
Manitoba	30,096	2266	13.28 (12.4)	3.80 (1.32)	—
Year of rating					
2005	138	132	1.05 (0.2)	3.75 (1.23)	138
2006	7726	4280	1.77 (1.4)	3.91 (1.25)	7588
2007	208,376	34,961	5.74 (5.3)	4.03 (1.23)	200,650
2008	293,001	28,945	2.92 (2.3)	3.86 (1.32)	84,625
2009	375,670	28,885	2.86 (2.2)	3.84 (1.33)	82,669
2010	454,689	30,384	2.60 (2.0)	3.82 (1.35)	79,019
2011	525,815	30,079	2.36 (1.8)	3.84 (1.33)	71,126
2012	588,803	29,436	2.14 (1.7)	3.84 (1.37)	62,988
2013	640,603	27,181	1.91 (1.3)	3.86 (1.42)	51,800

Figure 3. Proportion of mean ratings, per year, in the top 50th percentile of all rated physicians (2005-2013) with 95% confidence intervals depicted for each proportion.



Discussion

Using national-level data over a nearly 8-year period from the country's largest physician-rating website, we found that 57,412 unique physicians are rated online and that, overall, ratings are positive. We found differences in ratings with respect to physician specialty, geographic practice location, and year.

To our knowledge, this study is the first to describe the landscape of physician ratings in Canada. This adds to the body of national-level literature on physician-ratings websites in China, Germany, and the United States [6,10,12-14]. Previous studies have focused on either specific specialties or had shorter study periods [20-26]. Overall, our findings are in keeping with previous work that physician ratings are typically positive [6,12-14,17-19,28].

We found that certain specialties (eg, cardiac surgeons and nephrologists) were more likely to be rated in the top 50th percentile of all rated physicians, whereas others (eg, sleep disorder specialists, dermatologists, and addiction medicine specialists) were less likely to be rated as favorably. A variety of physician and patient factors may contribute to such differences. This may be due to differences in patient population as well as differences in patient expectations. For example, surviving a surgery may be a relatively straightforward "rateable" aspect for a surgeon; insight into recognizing the milestones for recovery from addiction with frequent relapses may not be as straightforward. In addition, there are likely more complex interactions between preconceived expectations patients have regarding their physician, their perceived performance of that physician, and their resulting satisfaction—as well described by the expectation-disconfirmation theory in the psychology and consumer marketing literature [42].

Our results add additional information and detail to previous work. Quality of ratings have been shown to be similar for physicians in primary care, medical specialties, surgeons and surgical specialties, and obstetrics and gynecology, but significantly differed for a category of "other physicians," which included radiologists, pathologists, and anesthesiologists [6]. Others have shown that pediatricians and surgeons had more favorable ratings, although others showed that ratings for generalists did not differ either in quantity or quality from those for subspecialists [17].

In addition to quality of ratings, we also looked at frequency of ratings by specialty. Certain specialties (eg, obstetrics, dermatology, and family medicine) were more commonly rated than others (ie, pathology and radiology), which based on their proportion in the national physician population, overall, in keeping with previous work [6,12,14]. One hypothesis is that patient-physician encounters during surgeries and pregnancies may be discrete care episodes that may be more amenable to appraisal. Also, specialties such as family medicine involve direct physician-patient interaction over time; in contrast, patients rarely interact with their pathologist or radiologist, the two least-rated specialties. Patients may also more readily attribute care to (and hence, rate) a single provider in the case of a surgeon, obstetrician, or dermatologist, as opposed to settings such as inpatient internal medicine, where multiple physicians may collaborate.

We also found differences in the likelihood of a positive rating for geographic location. It seems unlikely that physician quality vastly differs regionally, given the national accreditation and continuing education standards. We noted, in general, that east coast and territory provinces were more likely to have ratings greater than the median (4.5) compared with provinces west of Ontario. There may be geographic differences in rater expectations for a variety of reasons; for example, location may give rise to differences in accessibility to medical care. One interesting hypothesis is that when physicians are scarce, consumers may be more appreciative of access to a physician and this may bias their ratings in a more favorable manner. In addition, we looked at economic prosperity indicators such as gross domestic product by province and found that, overall, lower patient satisfaction is found in more economically prosperous provinces (ie, central and western provinces) [43], in contrast to a theory by Grigoroudis et al [44] that posits that higher patient satisfaction may be explained by economic prosperity. Moreover, other sociologic or cultural phenomenon across locations may lead to variable consumer preference, a well-described marketing phenomenon known as *geographic segmentation* [45]. Explanations for such differences are likely multifactorial and remain, as yet, unknown. There is limited research on the variability of online physician ratings with geographic practice location. Gao et al [6] reported that physicians in the southern United States were slightly more likely to be rated favorably than those practicing in the rest of the country. However, others have reported no difference in ratings regarding practice location and city size for certain surgical specialties [20-22].

Finally, we found differences for ratings over time. We suspect that this is due to patient factors, rather than physician factors, because we would not expect physician quality to fluctuate dramatically from year to year, and the survey instrument was consistent throughout the time period. Of note, RateMDs was founded in 2004 in the United States and, by 2005, online physician websites were still new in Canada (138 ratings in 2005). By 2007, popularity peaked at 200,650 ratings before stabilizing and decreasing by 2013. It is challenging to explain this phenomenon. It may be that in 2007, online physician ratings finally received public attention, resulting in a flood of "early adopters," which subsequently waned. There was sufficient popularity of such websites and several prominent nationwide media articles in 2007 that physicians became concerned about their use. One article in a popular national news source reported the Canadian Medical Association's displeasure at such sites and, in particular, warned of the potential for libel [46-48]. However, our findings suggest that these early users were actually more likely to post favorable ratings. This may be plausible, if only because physician-rating website users in general tend to have more positive views toward the Internet, despite no differences in total quantity of Internet usage from the general population [16]. This is in keeping with our finding that the likelihood of a positive rating was highest in 2007.

Since 2007, ratings stabilized and even decreased in absolute number through to 2013. This finding differs from US data, which shows physician-rating website usage rapidly increasing, although the study period in question spans a 5-year period that ends before this study making comparisons problematic [6]. Based on user traffic to competing physician websites in Canada, it does not appear that increasing popularity of competing websites is the explanation. Compared to the United States, in Canada there is comparatively less consumer choice in physician selection because the avenue to seek subspecialty consultation is via one's primary care physician rather than self-referral. This may, in turn, be driving a decrease in the popularity of physician-rating websites. This hypothesis has been used to explain the use of physician websites in England; although increasing over time as well, they have demonstrated a more gradual, stable rise in popularity compared to the rapidly accelerating US growth [39].

We acknowledge several limitations to our work. First, although our dataset spans nearly an 8-year period, we are missing data from a period of 3 months (ie, October-December 2013 to complete calendar year 2013). However, we feel a national database of greater than 57,000 physicians for nearly an 8-year period is sufficient to elucidate broad trends. Second, online physician-ratings data may not be generalizable. Rating website users likely differ from the general population by virtue of computer access and ability, and by their inclination to post ratings [30]. In addition, because all physicians are entered into the website by raters, it is possible that a physician may have two unique profiles. This database was deidentified; therefore, we were unable to ensure that duplicate profiles were corrected. Moreover, ratings are anonymously posted, so it is possible that fraudulent ratings exist; however, the website has quality control mechanisms in place to circumvent multiple fraudulent ratings (eg, deleting multiple reviews from a single Web address). Third, we could not control for the possibility that online ratings may, themselves, influence future ratings. For example, when a user logs onto the website to post a rating, their original inclination may be influenced by what has previously been published. Overall, these are issues that are germane to most physician-ratings websites and, on balance, we do not feel these limitations would significantly alter our observations, greatly affect broad trends of average ratings and regional differences, nor affect our conclusions.

This study provides new national-level information on the nature of online physician ratings, particularly regarding specialty, geographic practice location, and changes over time. It remains to be seen whether such trends will continue. The utility of online ratings for ascertaining and evaluating physician quality is still in question—and we would argue that before undertaking these larger questions, a better understanding of the scope and breadth of online physician ratings is required. Our study has shown important differences in how physicians are rated based on a physician's specialty, practice location, and the year in which the physician is rated. Further studies endeavor to better understand the scope, breadth, and utility of online physician ratings; in the meantime, what we do know is that such websites reflect the unsolicited views of the health care consumer and, as such, remain a valuable data source of the patient experience.

Authors' Contributions

All authors (JL, JM, CB) had access to the data and contributed to the conception or design of the work; the acquisition, analysis, or interpretation of data; the writing of the manuscript; gave final approval of the version to be published; and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of the work are appropriately investigated and resolved.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Proportion of ratings in top 50th percentile.

[PDF File (Adobe PDF File), 48KB - jmir_v20i3e76_app1.pdf]

References

1. Stelfox HT, Gandhi TK, Orav EJ, Gustafson M. The relation of patient satisfaction with complaints against physicians and malpractice lawsuits. Am J Med 2005;118:1126-1133. [Medline: 1614644]
2. Hickson GB, Federspiel CF, Pichert JW, Miller CS, Gauld-Jaeger J, Bost B. Patient complaints and malpractice risk. JAMA 2002;287:2951-2957. [Medline: 12052123]
3. Boulding W, Glickman SW, Manary MP, Schulman KA, Staelin R. Relationship between patient satisfaction with inpatient care and hospital readmission within 30 days. Am J Manag Care 2011 Jan;17(1):41-48 [FREE Full text] [Medline: 21348567]
4. Glickman SW, Boulding W, Manary M, Staelin R, Roe MT, Wolosin RJ, et al. Patient satisfaction and its relationship with clinical quality and inpatient mortality in acute myocardial infarction. Circ Cardiovasc Qual Outcomes 2010 Mar;3(2):188-195 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.109.900597] [Medline: 20179265]

5. 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]
6. 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]
7. McLennan S, Strech D, Reimann S. Developments in the frequency of ratings and evaluation tendencies: a review of German physician rating websites. J Med Internet Res 2017 Aug 25;19(8):e299 [FREE Full text] [doi: 10.2196/jmir.6599] [Medline: 28842391]
8. NHS Choices. Find hospital-consultants URL: https://www.nhs.uk/Service-Search/Hospital/LocationSearch/8/Consultants [accessed 2017-10-23] [WebCite Cache ID 6uQyHOl3V]
9. Weisse Liste. URL: https://www.weisse-liste.de/de/arzt/arztsuche/ [accessed 2017-10-23] [WebCite Cache ID 6uQygLHgi]
10. Hanauer DA, Zheng K, Singer DC, Gebremariam A, Davis MM. Public awareness, perception, and use of online physician rating sites. JAMA 2014 Feb 19;311(7):734-735. [doi: 10.1001/jama.2013.283194] [Medline: 24549555]
11. Emmert M, Sauter L, Jablonski L, Sander U, Taheri-Zadeh F. Do physicians respond to web-based patient ratings? An analysis of physicians' responses to more than one million web-based ratings over a six-year period. J Med Internet Res 2017 Jul 26;19(7):e275 [FREE Full text] [doi: 10.2196/jmir.7538] [Medline: 28747292]
12. 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 [FREE Full text] [doi: 10.2196/jmir.2655] [Medline: 23919987]
13. Emmert M, Sander U, Pisch F. Eight questions about physician-rating websites: a systematic review. J Med Internet Res 2013;15(2):e24 [FREE Full text] [doi: 10.2196/jmir.2360] [Medline: 23372115]
14. Hao H. The development of online doctor reviews in China: an analysis of the largest online doctor review website in China. J Med Internet Res 2015 Jun 01;17(6):e134 [FREE Full text] [doi: 10.2196/jmir.4365] [Medline: 26032933]
15. Merrell JG, Levy BH, Johnson DA. Patient assessments and online ratings of quality care: a “wake-up call” for providers. Am J Gastroenterol 2013 Nov;108(11):1676-1685. [doi: 10.1038/ajg.2013.112] [Medline: 24192941]
16. Terlutter R, Bidmon S, Röttl J. Who uses physician-rating websites? Differences in sociodemographic variables, psychographic variables, and health status of users and nonusers of physician-rating websites. J Med Internet Res 2014;16(3):e97 [FREE Full text] [doi: 10.2196/jmir.3145] [Medline: 24686918]
17. 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]
18. 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]
19. Emmert M, Meier F, Heider A, Dürr C, Sander U. What do patients say about their physicians? An analysis of 3000 narrative comments posted on a German physician rating website. Health Policy 2014 Oct;118(1):66-73. [doi: 10.1016/j.healthpol.2014.04.015] [Medline: 24836021]
20. Frost C, Mesfin A. Online reviews of orthopedic surgeons: an emerging trend. Orthopedics 2015 Apr;38(4):e257-e262. [doi: 10.3928/01477447-20150402-52] [Medline: 25901617]
21. Ellimoottil C, Hart A, Greco K, Quek ML, Farooq A. Online reviews of 500 urologists. J Urol 2013 Jun;189(6):2269-2273. [doi: 10.1016/j.juro.2012.12.013] [Medline: 23228385]
22. 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]
23. Sobin L, Goyal P. Trends of online ratings of otolaryngologists: what do your patients really think of you? JAMA Otolaryngol Head Neck Surg 2014 Jul;140(7):635-638. [doi: 10.1001/jamaoto.2014.818] [Medline: 24876073]
24. Gilbert K, Hawkins CM, Hughes DR, Patel K, Gogia N, Sekhar A, et al. Physician rating websites: do radiologists have an online presence? J Am Coll Radiol 2015 Aug;12(8):867-871. [doi: 10.1016/j.jacr.2015.03.039] [Medline: 26250980]
25. 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]
26. Trehan SK, Daluiski A. Online patient ratings: why they matter and what they mean. J Hand Surg Am 2016 Feb;41(2):316-319. [doi: 10.1016/j.jhsa.2015.04.018] [Medline: 26754195]
27. Hanauer DA, Zheng K, Singer DC, Gebremariam A, Davis MM. Parental awareness and use of online physician rating sites. Pediatrics 2014 Oct;134(4):e966-e975 [FREE Full text] [doi: 10.1542/peds.2014-0681] [Medline: 25246629]
28. 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]
29. Holliday AM, Kachalia A, Meyer GS, Sequist TD. Physician and patient views on public physician rating websites: a cross-sectional study. J Gen Intern Med 2017 Jun;32(6):626-631. [doi: 10.1007/s11606-017-3982-5] [Medline: 28150098]

30. Segal J, Sacopulos M, Sheets V, Thurston I, Brooks K, Puccia R. Online doctor reviews: do they track surgeon volume, a proxy for quality of care? J Med Internet Res 2012;14(2):e50 [FREE Full text] [doi: 10.2196/jmir.2005] [Medline: 22491423]
31. Gray BM, Vandergrift JL, Gao GG, McCullough JS, Lipner RS. Website ratings of physicians and their quality of care. JAMA Intern Med 2015 Feb;175(2):291-293. [doi: 10.1001/jamainternmed.2014.6291] [Medline: 25437252]
32. Adelhardt T, Emmert M, Sander U, Wambach V, Lindenthal J. Can patients rely on results of physician rating websites when selecting a physician? - A cross-sectional study assessing the association between online ratings and structural and quality of care measures from two German physician rating websites. Value Health 2015 Nov;18(7):A545 [FREE Full text] [doi: 10.1016/j.jval.2015.09.1732] [Medline: 26533058]
33. Liu JJ, Matelski J, Cram P, Urbach DR, Bell CM. Association between online physician ratings and cardiac surgery mortality. Circ Cardiovasc Qual Outcomes Nov 2016;9(6):788-791. [Medline: 29074628]
34. Okike K, Peter-Bibb TK, Xie KC, Okike ON. Association between physician online ratings and quality of care. J Med Internet Res 2016 Dec 13;18(12):e324 [FREE Full text] [doi: 10.2196/jmir.6612] [Medline: 27965191]
35. 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]
36. Murphy GP, Awad MA, Osterberg EC, Gaither TW, Chumnarnsongkhroh T, Washington SL, et al. Web-based physician ratings for California physicians on probation. J Med Internet Res 2017 Aug 22;19(8):e254 [FREE Full text] [doi: 10.2196/jmir.7488] [Medline: 28830852]
37. Daskivich TJ, Houman J, Fuller G, Black JT, Kim HL, Spiegel B. Online physician ratings fail to predict actual performance on measures of quality, value, and peer review. J Am Med Inform Assoc 2017 Sep 08:e1. [doi: 10.1093/jamia/ocx083] [Medline: 29025145]
38. 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]
39. Greaves F, Millett C. Consistently increasing numbers of online ratings of healthcare in England. J Med Internet Res 2012;14(3):e94 [FREE Full text] [doi: 10.2196/jmir.2157] [Medline: 22742977]
40. RateMDs. URL: https://www.ratemds.com/ [accessed 2017-02-08] [WebCite Cache ID 6830Waa8]
41. Canadian Institute for Health Information. URL: https://secure.cihi.ca/estore/productSeries.htm?pc=PCC476 [accessed 2017-02-08] [WebCite Cache ID 68IRuYpx]
42. Oliver RL. Effect of expectation and disconfirmation on postexposure product evaluations-an alternative interpretation? J Applied Psych 1977;62(4):480.
43. Statistics Canada. URL: http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/econ15-eng.htm [accessed 2017-11-06] [WebCite Cache ID 6umHDMYBS]
44. Grigoroudis E, Nikolopoulou G, Zopounidis C. Customer satisfaction barometers and economic development: an explorative ordinal regression analysis. Total Qual Manag Bus 2008 May;19(5):441-460. [doi: 10.1080/14783360802018095]
45. Kahle LR. The nine nations of North America and the value basis of geographic segmentation. J Marketing 1986 Apr;50(2):37-47. [doi: 10.2307/1251598]
46. Mackay B. RateMDs.com nets ire of Canadian physicians. CMAJ 2007 Mar 13;176(6):754 [FREE Full text] [doi: 10.1503/cmaj.070239] [Medline: 17353522]
47. CBC News Technology and Science. 2007 Mar 12. Doctor rating website stirs up physicians URL: http://www.cbc.ca/news/technology/doctor-rating-website-stirs-up-physicians-1.683440 [accessed 2017-02-08] [WebCite Cache ID 68Ip0z64]
48. Smart Canucks. RateMDs.com allows patients to rate and read about their doctors and dentists [thread] URL: http://forum.smartcanucks.ca/18448-ratemds-com-allows-patients-rate-read-about-their-doctors-dentists-canada/ [accessed 2017-02-08] [WebCite Cache ID 68JFDQ5m]

Abbreviations

CIHI: Canadian Institute of Health Information

Edited by G Eysenbach; submitted 10.02.17; peer-reviewed by M Emmert, S Bidmon; comments to author 26.03.17; revised version received 18.08.17; accepted 16.11.17; published 07.03.18.

Please cite as:
Liu JJ, Matelski JJ, Bell CM
Scope, Breadth, and Differences in Online Physician Ratings Related to Geography, Specialty, and Year: Observational Retrospective Study
J Med Internet Res 2018;20(3):e76
URL: http://www.jmir.org/2018/3/e76/
doi:10.2196/jmir.7475
PMID:29514775

Original Paper

How Online Quality Ratings Influence Patients' Choice of Medical Providers: Controlled Experimental Survey Study

Niam Yaraghi[1,2], PhD; Weiguang Wang[3], MSc; Guodong (Gordon) Gao[3], PhD; Ritu Agarwal[3], PhD

[1]Department of Operations and Information Management, University of Connecticut, Stamford, CT, United States

[2]Center for Technology Innovation, The Brookings Institution, Washington, DC, United States

[3]Department of Decision, Operations and Information Technologies, Robert H Smith School of Business, University of Maryland at College Park, College Park, MD, United States

Corresponding Author:
Niam Yaraghi, PhD
Center for Technology Innovation
The Brookings Institution
1755 Massachusetts Ave NW
Washington, DC, 20036
United States
Phone: 1 2027632073
Email: niam.yaraghi@uconn.edu

Abstract

Background: In recent years, the information environment for patients to learn about physician quality is being rapidly changed by Web-based ratings from both commercial and government efforts. However, little is known about how various types of Web-based ratings affect individuals' choice of physicians.

Objective: The objective of this research was to measure the relative importance of Web-based quality ratings from governmental and commercial agencies on individuals' choice of primary care physicians.

Methods: In a choice-based conjoint experiment conducted on a sample of 1000 Amazon Mechanical Turk users in October 2016, individuals were asked to choose their preferred primary care physician from pairs of physicians with different ratings in clinical and nonclinical aspects of care provided by governmental and commercial agencies.

Results: The relative log odds of choosing a physician increases by 1.31 (95% CI 1.26-1.37; *P*<.001) and 1.32 (95% CI 1.27-1.39; *P*<.001) units when the government clinical ratings and commercial nonclinical ratings move from 2 to 4 stars, respectively. The relative log odds of choosing a physician increases by 1.12 (95% CI 1.07-1.18; *P*<.001) units when the commercial clinical ratings move from 2 to 4 stars. The relative log odds of selecting a physician with 4 stars in nonclinical ratings provided by the government is 1.03 (95% CI 0.98-1.09; *P*<.001) units higher than a physician with 2 stars in this rating. The log odds of selecting a physician with 4 stars in nonclinical government ratings relative to a physician with 2 stars is 0.23 (95% CI 0.13-0.33; *P*<.001) units higher for females compared with males. Similar star increase in nonclinical commercial ratings increases the relative log odds of selecting the physician by female respondents by 0.15 (95% CI 0.04-0.26; *P*=.006) units.

Conclusions: Individuals perceive nonclinical ratings provided by commercial websites as important as clinical ratings provided by government websites when choosing a primary care physician. There are significant gender differences in how the ratings are used. More research is needed on whether patients are making the best use of different types of ratings, as well as the optimal allocation of resources in improving physician ratings from the government's perspective.

(J Med Internet Res 2018;20(3):e99) doi:10.2196/jmir.8986

KEYWORDS

quality of health care; health care evaluation mechanisms

Introduction

To improve quality, foster competition, promote transparency, and help patients make informed decisions, it is critical for patients to have access to reliable information and make cognizant choices about their medical providers [1,2]. In recent years, a concerted effort in the United States has been put in place to develop and publicly report quality measures of medical providers [3].

The Centers for Medicare and Medicaid Services (CMS) is the most prominent governmental agency in the United States that collects, aggregates, and reports quality measures of different aspects of medical care. Through initiatives such as Hospital Compare [4], CMS reports quality data on both clinical and nonclinical aspects of medical services offered by different providers. Surgical complications, infections, readmission, and death rates are examples of metrics that measure the clinical aspects of medical care. Surveys of patients' experiences, such as the Hospital Consumer Assessment of Healthcare Providers and Systems, capture metrics that measure nonclinical aspects of care. In parallel with CMS, private and commercial agencies such as Vitals [5], RateMDs [6], and ProPublica [7] also collect and report quality metrics on both clinical and nonclinical aspects of care. Recent research shows that although the ratings provided by commercial agencies may be inconsistent with each other [8], they are more comprehensive and cover a broader range of domains than what is included in ratings reported by CMS [9,10].

Ratings of health care providers are growing in importance and popularity [11-18], affecting both the revenue and the reputation of medical providers [19-22]. For example, when CMS released its quality metrics of nursing homes to the public, the market share of 1-star facilities decreased by 8%, whereas the market share of 5-star facilities increased by more than 6% [23]. Similar effects have also been documented for hospitals [24]. Although nonclinical ratings provided by commercial agencies are correlated with the conventional measures of patient experience as reported by governmental agencies [25,26], the relationship between patient reviews and medical outcomes is not clear. Some studies find that patient satisfaction reported as nonclinical ratings is not associated with clinical outcomes [27-32], whereas others report a strong association between these two types of ratings [33,34]. For a review of literature on the association between the social media reviews and the clinical quality outcomes, see Verhoef et al [35].

Despite the significant differences between the types (clinical and nonclinical) and the sources (governmental and commercial agencies) of ratings, variations in their relative significance for patient choice of medical providers are not known. The purpose of this research was to fill this gap by uncovering the relative importance of these ratings in the decision-making processes of different groups of patients.

Methods

Data Source

We used a primary dataset consisting of responses of 1000 individuals who were each paid 50 cents to participate in an online experiment through Amazon Mechanical Turk (AMT) in October 2016. These individuals were all master users of AMT and live in the United States. According to AMT, a user achieves a master distinction by consistently completing requests with a high degree of accuracy. Masters must continue to pass AMT's statistical monitoring to maintain their status [36].

Table 1 provides a comparison of demographics between the sample in this study and the US population. In contrast to the US population, our sample consisted of less affluent, but more educated, younger adults. Although, when compared with the US population, our sample of AMT users consisted of younger and more technologically savvy individuals, we relied on this sample to conduct our analysis for the following reasons. First, given the question posed in this research, the sample did not need to be representative of the US population and, instead, only had to represent individuals who used information resources available on the Internet. As this study compared the importance of two information resources that are exclusively Web-based, its sample also had to include the individuals who could use resources on the Web. Second, prior research shows that despite limitations, data that are gathered from "AMT samples are at least as reliable as those obtained via traditional methods. Overall, AMT can be used to obtain high-quality data inexpensively and rapidly" [37].

Study Design

To determine how ratings on different attributes affect individuals' evaluations of medical providers, we designed an experiment and conducted a choice-based conjoint analysis [38] as a rigorous method of eliciting preferences [39]. We describe the method below.

The combination of 2 categories (clinical and nonclinical) and 2 sources (governmental and commercial agencies) resulted in 4 different types of ratings: clinical ratings provided by a governmental agency, nonclinical ratings provided by a governmental agency, clinical ratings provided by a commercial agency, and nonclinical ratings provided by a commercial agency. In this research, we use "governmental agency" and "public agency" interchangeably. We assigned a high or low value to each type of rating, and thereby created 16 profiles of hypothetical physicians. In a 1-to-4-star rating system, to induce appropriate variation, we used 2 stars to indicate low ratings and 4 stars to indicate high ratings. Each profile represented a physician with different ratings on the 4 categories. These profiles were *balanced*, which means that each of the 2 levels (2 and 4 stars) in each of the 4 types of ratings appeared the same number of times in physician profiles. Using these 16 profiles, we then created 8 pairs of physicians such that the 4 types of ratings in each pair were *orthogonal* [40]. This ensured that any pair of levels from different rating types appeared the same number of times in the design. We used % *mktex* [41] macro in SAS software (version 9.4) to create the balanced and orthogonal design. Table 2 shows the 16 profiles in 8 pairs.

In a Web-based interface, we first provided respondents with a brief tutorial on different sources and types of ratings. Specifically, we described the public agency as "the department of Health and Human Services, which is a branch of the federal government" and the commercial agency as "websites such as Yelp, RateMDs, Healthgrades, Vitals, Zocdoc, and DoctorScorecard."

Table 1. Characteristics of 949 respondents and the US population.

Variable and class	Sample, n (%)	Percentage of US population[a] (%)	H_0: $P_{USA}-P_{Sample}=0$[b] (z value)
Education			
Advanced degree	114 (12.0)	10.38	−1.65[c]
Bachelor's degree	381 (40.2)	18.88	−16.74[d]
Associate's degree	100 (10.5)	5.28	−7.25[d]
Some college, no degree	231 (24.3)	19.42	−3.83[d]
Trade or technical school	30 (3.2)	4.08	1.43
Graduated high school	90 (9.5)	29.63	13.59[d]
Less than high school	3 (0.3)	12.33	11.25[d]
Income, US $			
150,000 or more	30 (3.2)	13.57	9.36[d]
125,000-149,999	25 (2.6)	5.42	3.8[d]
100,000-124,999	70 (7.4)	8.71	1.45
75,000-99,999	123 (13.0)	12.26	−0.66
50,000-74,999	227 (23.9)	16.96	−5.71[d]
35,000-49,999	170 (17.9)	12.92	−4.58[d]
25,000-34,999	136 (14.3)	9.39	−5.22[d]
Less than 25,000	168 (17.7)	20.77	2.33[e]
Race			
Asian	56 (5.9)	5.70	−0.27
Black	66 (7.0)	13.30	5.76[d]
Hawaiian	1 (0.1)	0.20	0.62
Hispanic	53 (6.0)	17.8	9.84[d]
Indian	16 (1.7)	1.30	−1.06
White	757 (79.8)	76.90	−2.1[e]
Marital status			
Divorced	74 (7.8)	9.80	2.07[e]
Married/Domestic partner	471(49.6)	51.87	1.38
Separated	8 (0.8)	2.09	2.69[d]
Single/Never married	390 (41.1)	32.25	−5.83[d]
Widowed	6 (0.6)	5.72	6.75[d]
Gender			
Female	548 (57.7)	50.80	−4.28[d]
Male	401 (42.3)	49.20	4.28[d]
Age			
Younger than 65 years	924 (97.4)	87.00	−9.49[d]
65 years and older	25 (2.6)	13.00	9.49[d]

[a]Authors' analysis of characteristics of experiment participants. Demographics of US population are calculated based on the data provided by the US Census Bureau.

[b]The null hypothesis that the percentage in sample is equal to that of the US population.

[c]$P<.10$.

[d]$P<.01$.

[e]$P<.05$.

Table 2. Physician profiles used in choice-based conjoint experiment. "gGvernment" indicates that a public agency provides the ratings, and "Commercial" indicates that a private organization provides the ratings. In the Web-based interface, the hypothetical physician profiles in each pair were shown side-by-side and respondents were asked to choose the physician they prefer. The sequence of the pairs and the attributes in each profile were generated randomly to ensure that the order of the presentation of rank of the attributes did not influence the respondent's choice. The values of 2 or 4 in the table, respectively, indicate a "2" or "4" star rating in the physician profiles provided to respondents in the Web-based experiment.

Pair number[a]	Government rating		Commercial rating	
	Clinical	Nonclinical	Clinical	Nonclinical
One	2; 4	4; 2	2; 4	4; 2
Two	2; 4	4; 2	4; 2	4; 2
Three	2; 4	2; 2	2; 2	4; 2
Four	4; 2	2; 4	2; 4	4; 2
Five	4; 2	2; 4	4; 2	4; 2
Six	4; 2	4; 2	2; 4	4; 2
Seven	2; 4	2; 4	4; 2	4; 2
Eight	2; 4	2; 4	2; 4	2; 4

Figure 1. Screenshot of the choice-based conjoint experiment.

We also distinguished clinical and nonclinical ratings and explained to the survey respondents that clinical ratings by the public agency were determined "based on official statistics on how often physicians provide care that research shows leads to the best results for patients" and nonclinical ratings by the public agency were determined based on "a national survey that asks patients about their experiences with staff, nurses, and doctors during a recent visit to the doctor." Similarly, we explained that clinical ratings provided by the commercial agency were determined by "the patient online reviews about how patients evaluate the medical expertise of the doctor" and nonclinical ratings provided by the commercial agency were created based on "patient online reviews about their experiences with staff, nurses, and doctors during a recent visit to the doctor." To assess if respondents correctly distinguished the differences between the types and the sources of ratings, at the end of the survey, we asked them to describe each type of the ratings in their own words. Our examination of their responses confirmed that all respondents had fully understood different ratings.

We then presented the 8 pairs of hypothetical profiles of physicians in a random sequence and asked respondents to choose the physician they prefer in each pair. A screenshot of 1 of the 8 comparison pairs is presented in Figure 1, which corresponds to the choices in pair Seven as shown in Table 1. To simulate a realistic decision-making scenario, we asked the respondents to imagine that they have moved to a new town and have to choose a new primary care physician based solely on the 4 types of ratings provided to them. This approach ensured that the choice of the respondents in our experiment was only driven by the ratings and was not confounded by any

other factor outside of our model, such as insurance coverage, location, or race of the physician [42,43].

Once respondents finished the evaluation of physicians in the 8 pairs, we asked them a series of questions designed to evaluate their health status, medical literacy, trust in Web-based reviews, and trust in government as 4 composite indexes. We conducted factor analysis to operationalize these 4 constructs using validated items that we derived from prior literature in information systems [44,45] and medicine [46,47]. Details on the items, composite indexes, and factor analysis are provided in Multimedia Appendix 1.

One potential concern with the study design was that respondents may not complete the choice task thoughtfully. To detect and filter the responses that were provided hastily and without careful attention, we included 2 trap questions in the experiment.

The first trap question was the choice of physicians in the eighth pair (shown in Table 2), one of which was superior on all of the 4 types of ratings and clearly dominated the pair. A respondent's choice of an inferior physician indicated lack of attention to the experiment. The second trap question asked, "How happy will you be if you receive a letter from Internal Revenue Service that says you should pay a large amount of taxes to the government?" We assumed that a respondent did not pay attention to the question if she chose "extremely happy" or "happy" as a response to this question.

Statistical Analysis

Our research design fit the multinomial logit model with clustered error terms [48,49]. Following the suggestions of Kuhfeld [50], we used the PHREG [51] procedure in SAS software for the estimation. In this model, the dependent variable was binary and indicated the choice that a respondent made from a pair of hypothetical physician profiles. The 4 types of ratings in each profile constituted our main independent variables. In the multinomial logit model used in this study, the probability that a respondent chose a specific physician in a pair was a function of the attributes of that specific physicians as well as the attributes of the other physician in the pair. The PHREG [51] procedure in SAS not only allowed us to account for the conditional dependency of choices for the alternatives in a pair but also adjusted for the correlation between the 8 choices made by the same respondent. Using this model, we could examine the relative importance of the 4 types of ratings. We further explored whether patient attributes, such as age, gender, and income, moderated the impact of the ratings. To statistically compare the effects of different regression coefficients, we implemented the tests provided by Paternoster et al [52].

Results

On the basis of the answers to the 2 trap questions, we excluded 51 observations from our initial sample of 1000 responses. We retained the remaining 949 responses for further analysis (Table 1). We present the estimation results of our multinomial logit model in Table 3.

As shown in the last (full model) column of Table 3, the relative log odds of choosing a physician increased by 1.31 (95% CI 1.26-1.37; $P<.001$) and 1.32 (95% CI 1.27-1.39; $P<.001$) units when the government clinical ratings and commercial nonclinical ratings moved from 2 to 4 stars, respectively. The importance of these 2 types of ratings was statistically equivalent ($P=.49$). By comparison, the relative log odds of choosing a physician increased by a modest 1.12 (95% CI 1.07-1.18; $P<.001$) units when the commercial clinical ratings moved from 2 to 4 stars. The relative log odds of selecting a physician with 4 stars in nonclinical ratings provided by the government was 1.03 (95% CI 0.98-1.09; $P<.001$) units higher than a physician with 2 stars in this rating. The difference between the effects of government nonclinical ratings and commercial clinical ratings on patients' choice of a primary care physician were statically significant ($P=.04$). The difference between the effects of clinical ratings provided by government and those provided by a commercial agency was statistically significant ($P<.001$). Likewise, the difference between the government clinical ratings and the government nonclinical ratings was also statistically significant ($P<.001$).

One standard deviation improvement in a patient's health status increased the relative log odds of choosing a physician with 4 stars in commercial nonclinical ratings by 0.18 (95% CI 0.13-0.24; $P<.001$) units and decreased the relative log odds of choosing a physician with 4 stars in government clinical ratings by 0.14 (95% CI 0.08-0.19; $P<.001$) units.

Medical literacy had no statistically significant effect on how patients evaluated different types of ratings. As the level of trust in overall Web-based ratings increased, the importance of nonclinical ratings provided by a commercial agency also increased. One standard deviation increase in a patient's trust in Web-based reviews increased the relative log odds of choosing a physician with 4 stars in nonclinical commercial ratings by 0.07 (95% CI 0.02-0.13; $P=.05$) units. Unsurprisingly, as the patients' level of trust in the government increased, the importance of clinical ratings provided by government increased, whereas the importance of nonclinical ratings provided by a commercial agency decreased.

One standard deviation increase in a patient's trust in government increased the relative log odds of choosing a physician with 4 stars in government clinical ratings by 0.20 (95% CI 0.15-0.25; $P<.001$) units and decreased the relative log odds of choosing a physician with 4 stars in commercial nonclinical ratings by −0.15 (95% CI 0.10-0.21; $P<.001$) units. These trends remained consistent even when we included more variables in our model. We also examined how patients' demographic characteristics of gender, race, income, education, marital status, and age affected the importance of each of the 4 ratings in their evaluation of primary care physicians. Table 4 presents the results. The log odds of selecting a physician with 4 stars in nonclinical government ratings relative to a physician with 2 stars was 0.23 (95% CI 0.13-0.33; $P<.001$) units higher for females compared with males. Similar star increase in nonclinical commercial ratings increased the relative log odds of selecting the physician by female patients by an additional 0.15 (95% CI 0.04-0.25; $P=.006$) units, compared with males.

Table 3. The relative importance of different types and sources of ratings on patients' choice. GC: clinical ratings provided by a public agency (government). GNC: nonclinical ratings provided by a public agency (government). YC: clinical ratings provided by a commercial agency (commercial). YNC: nonclinical ratings provided by a commercial agency (commercial).

Parameter[a]	Parameter estimate (95% CI)					
	Basic model	Health status	Medical literacy	Trust in online reviews	Trust in government	Full model
GC	1.29[b] (1.24 to 1.34)	1.29[b] (1.24 to 1.35)	1.29[b] (1.24 to 1.34)	1.29[b] (1.24 to 1.34)	1.30[b] (1.25 to 1.35)	1.31[b] (1.26 to 1.36)
GNC	1.00[b] (0.95 to 1.05)	1.01[b] (0.96 to 1.06)	1.00[b] (0.95 to 1.05)	1.00[b] (0.95 to 1.05)	1.01[b] (0.96 to 1.06)	1.03[b] (0.98 to 1.08)
YC	1.09[b] (1.04 to 1.14)	1.11[b] (1.06 to 1.16)	1.10[b] (1.04 to 1.14)	1.10[b] (1.04 to 1.14)	1.11[b] (1.07 to 1.16)	1.12[b] (1.07 to 1.18)
YNC	1.29[b] (1.24 to 1.34)	1.31[b] (1.25 to 1.36)	1.29[b] (1.24 to 1.34)	1.29[b] (1.24 to 1.35)	1.30[b] (1.25 to 1.35)	1.32[b] (1.27 to 1.37)
Health status × GC		−0.13[c] (−0.18 to −0.08)				−0.13[c] (−0.19 to −0.08)
Health status × GNC		0.09[c] (0.04 to 0.14)				0.10[c] (0.05 to 0.15)
Health status × YC		0.05 (0 to 0.10)				0.05 (0 to 0.10)
Health status × YNC		0.17[b] (0.12 to 0.22)				0.18[b] (0.13 to 0.23)
Medical literacy × GC			0 (−0.06 to 0.04)			0 (−0.06 to 0.04)
Medical literacy × GNC			0 (−0.05 to 0.04)			−0.01 (−0.05 to 0.04)
Medical literacy × YC			0.03 (−0.01 to 0.08)			0.03 (−0.01 to 0.08)
Medical literacy × YNC			−0.01 (−0.07 to 0.03)			−0.01 (−0.07 to 0.03)
Online trust × GC				0.06[d] (0.01 to 0.11)		0.06 (0.01 to 0.11)
Online trust × GNC				0.02 (−0.02 to 0.07)		0.02 (−0.02 to 0.07)
Online trust × YC				−0.05 (−0.10 to 0)		−0.05 (−0.10 to −0.01)
Online trust × YNC				0.07[d] (0.02 to 0.12)		0.07[d] (0.02 to 0.12)
Trust in government × GC					0.19[b] (0.14 to 0.24)	0.20[b] (0.14 to 0.25)
Trust in government × GNC					0.01 (−0.04 to 0.05)	0.01 (−0.04 to 0.05)
Trust in government × YC					0.03 (−0.02 to 0.08)	0.02 (−0.02 to 0.08)
Trust in government × YNC					−0.14[b] (−0.20 to −0.09)	−0.15[b] (−0.20 to −0.10)

[a]Authors' analysis of revealed choices in the choice-based conjoint analysis. Health status, medical literacy, online trust, and trust in government are composite indexes, centered around mean 0 with standard deviation of 1; 95% CI are reported in parentheses.

[b]*P*<.001.

[c]*P*<.01.

[d]*P*<.05.

Table 4. Interaction of ratings and patient characteristics. GC: clinical ratings provided by a public agency (government). GNC: nonclinical ratings provided by a public agency (government). YC: for clinical ratings provided by a commercial agency (commercial). YNC: the nonclinical ratings provided by a commercial agency (commercial).

Parameter	Parameter estimate (SE)						
	Female	White	High income	High education	Married	Age	Full model
GC	1.31[a] (0.05)	1.24[a] (0.08)	1.25[a] (0.05)	1.18[a] (0.05)	1.26[a] (0.05)	1.15[a] (0.12)	1.03[a] (0.14)
GNC	0.87[a] (0.05)	1.16[a] (0.08)	0.98[a] (0.04)	1.03[a] (0.05)	0.97[a] (0.04)	1.22[a] (0.11)	1.25[a] (0.14)
YC	1.15[a] (0.05)	1.15[a] (0.08)	1.08[a] (0.05)	1.11[a] (0.05)	1.07[a] (0.04)	1.32[a] (0.11)	1.42[a] (0.14)
YNC	1.19[a] (0.05)	1.30[a] (0.08)	1.24[a] (0.05)	1.29[a] (0.05)	1.18[a] (0.05)	1.21[a] (0.12)	1.13[a] (0.14)
GC × Female	−0.03 (0.05)						−0.04 (0.05)
GNC × Female	0.23[a] (0.05)						0.23[a] (0.05)
YC × Female	−0.09 (0.05)						−0.10 (0.05)
YNC × Female	0.18[b] (0.05)						0.15[b] (0.05)
GC × White		0.05 (0.09)					0.03 (0.10)
GNC × White		−0.19[c] (0.09)					−0.17 (0.09)
YC × White		−0.08 (0.09)					−0.05 (0.09)
YNC × White		−0.01 (0.09)					−0.04 (0.10)
GC × Income			0.074 (0.07)				0.01 (0.08)
GNC × Income			0.02 (0.07)				0.02 (0.07)
YC × Income			0.01 (0.07)				0 (0.07)
YNC × Income			0.10 (0.07)				0.03 (0.08)
GC × Education				0.21[b] (0.07)			0.21[b] (0.07)
GNC × Education				−0.06 (0.07)			−0.07 (0.07)
YC × Education				−0.04 (0.07)			−0.04 (0.07)
YNC × Education				0 (0.07)			−0.01 (0.07)
GC × Married					0.05 (0.07)		0.03 (0.08)
GNC × Married					0.05 (0.07)		0.046 (0.07)
YC × Married					0.04 (0.07)		0.09 (0.08)
YNC × Married					0.22[b] (0.07)		0.18[c] (0.08)
GC × Age						0.003 (0.003)	0.01 (0.01)
GNC × Age						−0.006[c] (0.002)	−0.006[c] (0.003)
YC × Age						−0.006[c] (0.002)	−0.006[c] (0.003)
YNC × Age						0.002 (0.003)	0 (0.003)

[a]*P*<.001.

[b]*P*<.01.

[c]*P*<.05.

Discussion

Principal Findings

To the best of our knowledge, this was the first research that, using a conjoint analysis, uncovered how individuals used Web-based ratings to compare and choose medical providers. We found that the clinical ratings provided by the government and the nonclinical ratings provided by a commercial agency were significantly more important for patient choice than nonclinical ratings provided by the government or clinical ratings provided by commercial agencies. We also found some differences in the importance of ratings based on the sociodemographic and health characteristics of respondents. Healthier patients paid more attention to nonclinical ratings, especially those from a commercial agency. On the other hand, for healthier patients, the importance of clinical ratings, notably those that are provided by the government, was lower. We found that female patients gave more importance to nonclinical ratings provided by both public and commercial agencies, compared with males. In comparison with other races, white respondents paid less attention to the nonclinical ratings provided by government. There was no other difference between racial groups in the importance of different types of ratings in the physician choice decision. Income did not play a role in the way respondents used the ratings in their decision. As patients get older, nonclinical ratings provided by the government and the clinical ratings provided by a commercial agency became even less important in how they evaluated medical providers.

A particular strength of this study was that we utilized a carefully controlled experimental design to observe the revealed preferences of participants rather than merely asking them to state them in response to a questioner, which could otherwise be subject to attribution or social desirability biases. Revealed preferences elicited in this experiment provided a more natural context, even when presented in hypothetical settings, and gave us greater confidence that the effects we observed within the sample were driven by the conjoint attributes rather than other unobserved factors.

Limitations

One limitation of our study was that we rated the attributes of the physicians by either 2 or 4 stars, whereas in reality, the ratings usually have 5 levels, between 1 and 5 stars. We limited the ratings to only 2 levels to reduce the number of possible combinations. If we considered 5 levels for each rating, the number of possible physician profiles would have surged from 16 to 625. Respondents could not reasonably compare these many physician profiles with each other. A second limitation of this study was that, in comparison with the US general population, its sample was drawn from younger, more educated, and less affluent individuals. Although samples from AMT have been shown to respond similarly to representative samples of the US population [37], the results from the study must be interpreted in light of the characteristics of the sample. Third, this study only focused on American respondents, and therefore, findings may not generalize to individuals outside of the United States. This was due to the fact that constructs such as medical literacy, health status, and trust in government significantly vary across individuals from different countries. Moreover, the presence of commercial websites and the availability of alternative government websites also vary across countries, which represents a further limitation on generalizability. Finally, in our study, we did not ask respondents whether they were familiar with the sources of information they were being asked to evaluate, primarily because our major focus was on the source (ie, government vs commercial) rather than a specific website. Future experiments could also ask respondents about their familiarity with the sources of information that they are asked to evaluate in the experiment.

Future Research

There are 3 potential areas for further research. The first is to examine how familiar individuals are with the sources of information provided by governmental and commercial agencies. Although most individuals are now fairly familiar with the commercial rating websites, knowledge about the other sources of information provided through governmental websites may be limited. It would be useful to quantify the level of awareness of such information as a precursor to designing appropriate policies to inform the public. The second is to replicate this study on an international sample to investigate how individuals outside of the United States rely on different sources and types of information for choosing their primary care physicians. Finally, the relative importance of Web-based ratings in comparison with other factors such as insurance coverage, recommendations of family and friends, and proximity to patients' residence is still unclear and could be investigated in future research.

Policy Recommendations

The findings of this research have implications for policy makers and medical providers. Although the government has expended substantial resources on clinical quality ratings, our study indicates a need to also acknowledge the importance of nonclinical measures. This is consistent with the recent CMS efforts and policy recommendations [53] to tie reimbursements to patient satisfaction. To the extent that nonclinical ratings appear to be more important for healthier patients, it clearly underscores the important role played by the "experience" of interacting with a physician for individuals whose visits to the doctor are likely to be preventive rather than curative. Primary care providers can consider ways in which the patient's experience can be improved, such as reduced waiting time and more empathetic interactions, which will eventually be reflected in the nonclinical ratings they receive. The results of this study could also encourage a public relations campaign to increase public awareness of the reviews that are government maintained and are more clinically based. Our result on gender differences in the relative salience of nonclinical ratings further revealed the importance of improving the patient experience for providers who are focused on women's health services.

With respect to patients' age, we found that older patients and those who trusted government more paid more attention to government-provided ratings. This is corroborated by prior literature, which documents that citizens who trust government more are also more satisfied with government websites [54]. We therefore recommend that CMS create website content and

user experiences that are tailored for Medicare beneficiaries and older patients as they rely on government-provided information more than the younger patients. Our results also indicated that commercial websites can be more successful in attracting younger individuals. If CMS intends to expand its audience, it should consider information dissemination strategies that appeal to patients in this segment.

Given the recent apprehensions expressed about the quality and representativeness of ratings provided by commercial websites [55], it is a matter of some concern that patients gave equal importance to commercial ratings of nonclinical aspects of care much as they did to government ratings of clinical aspects of care. This is likely a result of the richness of the information that patients believe they can receive from other patients who have engaged in interactions with the medical provider. It might also be driven by other factors such as the first mover advantage of commercial organizations as they have been active in rating a wide variety of services earlier than other governmental agencies. To that end, our findings suggest that patients have developed a preference for commercial websites for experience-based ratings of medical providers, that is, ratings that primarily capture information about the patient's experience with the medical provider. Thus, government agencies that offer similar ratings should pay careful attention to improve the usability of the information while concurrently addressing any perceptual obstacles that may prevent consumers from using these ratings.

Conclusions

Our research shows that patients pay equal attention to both clinical and nonclinical ratings when choosing a primary care physician. To obtain information about clinical ratings, they rely more on government sources, whereas for information on nonclinical ratings, they rely more on commercial sources. Both public and private agencies expend significant resources to design metrics, collect data, calculate ratings, and report them to the public. These resources are limited and should be optimally allocated to the type of ratings that consumers appreciate and will use the most. The findings of this research highlight the importance of efforts from government agencies such as CMS to improve its reporting of nonclinical ratings. Given the importance of nonclinical ratings in patients' decision making, we recommend that medical providers pay close attention to their nonclinical ratings on commercial websites as they represent a consequential source of customer feedback for improving the patient experience. Ultimately, the overarching objective of all rating sources must be focused on protecting patients from incorrect or misleading data, while simultaneously educating them on how best to interpret and make best use of the information presented.

Acknowledgments

GG and WW are partially supported by NSF Career Award #1254021.

Conflicts of Interest

None declared.

Multimedia Appendix 1

How online quality ratings influence patients' choice of medical providers: a controlled experimental survey study appendix (online-only material).

[PDF File (Adobe PDF File), 69KB - jmir_v20i3e99_app1.pdf]

References

1. Yegian J, Dardess P, Shannon M, Carman K. Engaged patients will need comparative physician-level quality data and information about their out-of-pocket costs. Health Aff (Millwood) 2013;32(2):328-337. [Medline: 23381526]
2. Hibbard J, Slovic P, Peters E, Finucane M. Strategies for reporting health plan performance information to consumers: evidence from controlled studies. Health Serv Res 2002;37(2):291-313. [doi: 10.1111/1475-6773.024] [Medline: 12035995]
3. James J, Felt-Lisk S, Werner R, Agres T, Schwartz A, Dentzer S. Issuelab. 2012. Health policy brief: public reporting on quality and costs URL: https://www.issuelab.org/resource/health-affairs-rwjf-health-policy-brief-public-reporting-on-quality-and-costs.html [accessed 2018-02-20] [WebCite Cache ID 6xMlzwxjb]
4. Medicare. Hospital compare URL: https://www.medicare.gov/hospitalcompare/search.html? [accessed 2018-02-16] [WebCite Cache ID 6xHEg8zCo]
5. Vitals. URL: http://www.vitals.com/ [accessed 2018-02-16] [WebCite Cache ID 6xHEhYlON]
6. RateMDs. Doctor reviews and ratings near you URL: https://www.ratemds.com/dc/washington/ [accessed 2017-02-13] [WebCite Cache ID RateMDs]
7. Wei S, Pierce O, Allen M. ProPublica. Surgeon scorecard URL: https://projects.propublica.org/surgeons/ [accessed 2018-02-16] [WebCite Cache ID 6xHEjXAnt]
8. Austin J, Jha A, Romano P, Singer S, Vogus T, Wachter R, et al. National hospital ratings systems share few common scores and may generate confusion instead of clarity. Health Aff (Millwood) 2015;34(3):423-430. [Medline: 25732492]

9. Ranard B, Werner R, Antanavicius T, Schwartz H, Smith R, Meisel Z, et al. Yelp reviews of hospital care can supplement and inform traditional surveys of the patient experience of care. Health Aff (Millwood) 2016;35(4):697-705. [Medline: 27044971]
10. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Harnessing the cloud of patient experience: using social media to detect poor quality healthcare. BMJ Qual Saf 2013;22(3):4. [doi: 10.1136/bmjqs-2012-001527]
11. 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(1):332. [doi: 10.1186/1472-6963-10-332]
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 Feb;14(1):e38 [FREE Full text] [doi: 10.2196/jmir.2003] [Medline: 22366336]
13. Hanauer D, Zheng K, Singer D, Gebremariam A, Davis M. Public awareness, perception, and use of online physician rating sites. JAMA 2014;311(7):734-735. [Medline: 24549555]
14. Greaves F, Millett C. Consistently increasing numbers of online ratings of healthcare in England. J Med Internet Res 2012 Jun 29;14(3):e94 [FREE Full text] [doi: 10.2196/jmir.2157] [Medline: 22742977]
15. Findlay S. Consumers' interest in provider ratings grows, and improved report cards and other steps could accelerate their use. Health Aff (Millwood) ? 2016;35(4):688-696. [Medline: 27044970]
16. 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. [Medline: 28842391]
17. Zwijnenberg N, Hendriks M, Bloemendal E, Damman O, de JJ, Delnoij D, et al. Patients' need for tailored comparative health care information: a qualitative study on choosing a hospital. J Med Internet Res 2016 Nov 28;18(11):e297 [FREE Full text] [doi: 10.2196/jmir.4436] [Medline: 27895006]
18. 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]
19. Sobin L, Goyal P. Trends of online ratings of otolaryngologists: what do your patients really think of you? JAMA Otolaryngol Head Neck Surg 2014 Jul;140(7):635-638. [doi: 10.1001/jamaoto.2014.818] [Medline: 24876073]
20. Mehta SJ. Patient satisfaction reporting and its implications for patient care. AMA J Ethics 2015 Jul 01;17(7):616-621 [FREE Full text] [doi: 10.1001/journalofethics.2015.17.7.ecas3-1507] [Medline: 26158808]
21. Bakhsh W, Mesfin A. Online ratings of orthopedic surgeons: analysis of 2185 reviews. Am J Orthop (Belle Mead NJ) 2014 Aug;43(8):359-363. [Medline: 25136868]
22. Elliott M, Beckett M, Lehrman W, Cleary P, Cohea C, Giordano L, et al. Understanding the role played by medicare's patient experience points system in hospital reimbursement. Health Aff (Millwood) 2016;35(9):1680. [doi: 10.1377/hlthaff.2015.0691]
23. Werner RM, Konetzka RT, Polsky D. Changes in consumer demand following public reporting of summary quality ratings: an evaluation in nursing homes. Health Serv Res 2016 Jun;51 Suppl 2:1291-1309. [doi: 10.1111/1475-6773.12459] [Medline: 26868034]
24. Chandra A, Finkelstein A, Sacarny A, Syverson C. Health care exceptionalism? performance and allocation in the us health care sector. Am Econ Rev 2016 Aug;106(8):2110-2144 [FREE Full text] [doi: 10.1257/aer.20151080] [Medline: 27784907]
25. 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]
26. Greaves F, Pape U, King D, Darzi A, Majeed A, Wachter R, et al. Associations between internet-based patient ratings and conventional surveys of patient experience in the English NHS: an observational study. BMJ Qual Saf 2012;21(7):605. [Medline: 22523318]
27. Franks P, Fiscella K, Shields C, Meldrum S, Duberstein P, Jerant A, et al. Are patients' ratings of their physicians related to health outcomes? Ann Fam Med 2005;3(3):234. [Medline: 15928226]
28. Gray BM, Vandergrift JL, Gao GG, McCullough JS, Lipner RS. Website ratings of physicians and their quality of care. JAMA Intern Med 2015 Feb;175(2):291-293. [doi: 10.1001/jamainternmed.2014.6291] [Medline: 25437252]
29. Campbell L, Li Y. Are Facebook user ratings associated with hospital cost, quality and patient satisfaction? a cross-sectional analysis of hospitals in New York state. BMJ Qual Saf 2018;27(2):119-129. [Medline: 28600452]
30. Okike K, Peter-Bibb T, Xie K, Okike O. Association between physician online rating and quality of care. J Med Internet Res 2016;18(12):e324. [doi: 10.2196/jmir.6612]
31. Greaves F, Pape U, King D, Darzi A, Majeed A, Wachter R, et al. Associations between web-based patient ratings and objective measures of hospital quality. Arch Intern Med 2012;172(5):435-436. [Medline: 22331980]
32. Greaves F, Pape U, Lee H, Smith D, Darzi A, Majeed A, et al. Patients' ratings of family physician practices on the internet: usage and associations with conventional measures of quality in the English National health service. J Med Internet Res 2012;14(5):e146. [Medline: 23076301]

33. Glickman SW, Boulding W, Manary M, Staelin R, Roe MT, Wolosin RJ, et al. Patient satisfaction and its relationship with clinical quality and inpatient mortality in acute myocardial infarction. Circ Cardiovasc Qual Outcomes 2010 Mar;3(2):188-195 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.109.900597] [Medline: 20179265]
34. 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]
35. Verhoef L, Van de Belt T, Engelen L, Schoonhoven L, Kool R. Social media and rating sites as tools to understanding quality of care: a scoping review. J Med Internet Res 2014;16(2):e56. [Medline: 24566844]
36. Mturk. Amazon mechanical turk URL: https://www.mturk.com/worker/help [accessed 2018-02-16] [WebCite Cache ID 6xHFQPi9I]
37. Buhrmester M, Kwang T, Gosling S. Amazon's mechanical turk: a new source of inexpensive, yet high-quality, data? Perspect Psychol Sci 2011;6(1):3-5. [doi: 10.1177/1745691610393980]
38. Green PE, Srinivasan V. Conjoint analysis in marketing: new developments with implications for research and practice. J Mark 1990;54(4):19.
39. Ryan M, Farrar S. Using conjoint analysis to elicit preferences for health care. BMJ 2000 Jun 03;320(7248):1530-1533 [FREE Full text] [Medline: 10834905]
40. Hauser J. MIT. 2007. A note on conjoint analysis URL: http://www.mit.edu/~hauser/Papers/NoteonConjointAnalysis.pdf [accessed 2018-02-16] [WebCite Cache ID 6xHFiVCmd]
41. Kuhfeld W. SAS. Orthogonal arrays URL: http://support.sas.com/techsup/technote/ts723.html [accessed 2018-02-16] [WebCite Cache ID 6xHFixRqX]
42. Saha S, Taggart SH, Komaromy M, Bindman AB. Do patients choose physicians of their own race? Health Aff (Millwood) 2000;19(4):76-83 [FREE Full text] [Medline: 10916962]
43. Blizzard R. Gallup. Healthcare Panel: How Do People Choose Hospitals? Internet. Gallup; 2005. Gallup news URL: http://news.gallup.com/poll/19402/Healthcare-Panel-How-People-Choose-Hospitals.aspx [accessed 2018-02-16] [WebCite Cache ID 6xHFoPbwG]
44. Bonelli S. Bright Local. 2016. Local consumer review survey 2016 URL: https://www.brightlocal.com/learn/local-consumer-review-survey/ [accessed 2018-02-16] [WebCite Cache ID 6xHG4knqW]
45. McKnight D, Choudhury V, Kacmar C. Developing and validating trust measures for e-commerce: an integrative typology. Inf Syst Res 2002;13(3):334-359.
46. Wallston K, Cawthon C, McNaughton C, Rothman R, Osborn C, Kripalani S. Psychometric properties of the brief health literacy screen in clinical practice. J Gen Intern Med 2014;29(1):119-126. [Medline: 23918160]
47. Chew L, Bradley K, Boyko E. Brief questions to identify patients with inadequate health literacy. Fam Med 2004;36(8):588-594. [Medline: 15343421]
48. McFadden D. Berkeley. Front Econom Internet New York: Academic Press; 1973. Conditional logit analysis of qualitative choice behavior URL: https://elsa.berkeley.edu/reprints/mcfadden/zarembka.pdf [accessed 2018-02-16] [WebCite Cache ID 6xHGN5CCd]
49. Manski C, McFadden D. Berkeley. Cambridge: MIT Press; 1981. Structural analysis of discrete data with econometric applications URL: https://emlab.berkeley.edu/~mcfadden/discrete/front_matter.pdf [accessed 2018-02-16] [WebCite Cache ID 6xHGNrMAS]
50. Kuhfeld W. SAS. Carry, NC: SAS; 2005. Marketing research methods in SAS URL: https://support.sas.com/techsup/technote/mr2010.pdf [accessed 2018-02-20] [WebCite Cache ID 6xMthtBFi]
51. SAS. Cary, NC: SAS Institute; 2008. The PHREG procedure URL: https://support.sas.com/documentation/cdl/en/statugphreg/61816/PDF/default/statugphreg.pdf [accessed 2018-02-20] [WebCite Cache ID 6xMtm2LSa]
52. Paternoster R, Brame R, Mazerolle P, Piquero A. Using the correct statistical test for the equality of regression coefficients. Criminology 1998;36(4):859-866. [Medline: 10.1111/j.1745-9125.1998.tb01268.x]
53. Jha A. JAMA. Payment power to the patients URL: https://newsatjama.jama.com/2017/05/22/jama-forum-payment-power-to-the-patients/ [accessed 2018-02-16] [WebCite Cache ID 6xHGYrIQB]
54. Welch E, Hinnant C, Moon M. Linking citizen satisfaction with e-government and trust in government. J Public Adm Res Theory 2005;15(3):371-391. [doi: 10.1093/jopart/mui021]
55. Lagu T, Metayer K, Moran M, Ortiz L, Priya A, Goff SL, et al. Website characteristics and physician reviews on commercial physician-rating websites. JAMA 2017 Feb 21;317(7):766. [doi: 10.1001/jama.2016.18553]

Abbreviations

AMT: Amazon Mechanical Turk
CMS: Centers for Medicare and Medicaid Services

Edited by G Eysenbach; submitted 16.09.17; peer-reviewed by H Miller, T Kool, E Leas; comments to author 12.10.17; revised version received 21.11.17; accepted 10.12.17; published 26.03.18.

Please cite as:
Yaraghi N, Wang W, Gao G(, Agarwal R
How Online Quality Ratings Influence Patients' Choice of Medical Providers: Controlled Experimental Survey Study
J Med Internet Res 2018;20(3):e99
URL: http://www.jmir.org/2018/3/e99/
doi:10.2196/jmir.8986
PMID:29581091

Original Paper

Patient Centeredness in Electronic Communication: Evaluation of Patient-to-Health Care Team Secure Messaging

Timothy P Hogan[1,2,3], PhD; Tana M Luger[1,2,4], MPH, PhD; Julie E Volkman[3,5], PhD; Mary Rocheleau[1,2], MS; Nora Mueller[1,2], MA; Anna M Barker[1,2], MS; Kim M Nazi[6], PhD; Thomas K Houston[1,2,3], MPH, MD; Barbara G Bokhour[1,2,7], PhD

[1]Center for Evaluating Patient-Centered Care in the Veterans Health Administration, Edith Nourse Rogers Memorial Veterans Hospital, Bedford, MA, United States

[2]Center for Healthcare Organization and Implementation Research, Edith Nourse Rogers Memorial Veterans Hospital, Bedford, MA, United States

[3]Division of Health Informatics and Implementation Science, Department of Quantitative Health Sciences, University of Massachusetts Medical School, Worcester, MA, United States

[4]Psychology Field Group, Pitzer College, Claremont, CA, United States

[5]Department of Communication, Bryant University, Smithfield, RI, United States

[6]Veterans and Consumers Health Informatics Office, Office of Connected Care, Veterans Health Administration, US Department of Veterans Affairs, Washington, DC, United States

[7]Department of Health, Law, Policy & Management, School of Public Health, Boston University, Boston, MA, United States

Corresponding Author:
Timothy P Hogan, PhD
Center for Healthcare Organization and Implementation Research
Edith Nourse Rogers Memorial Veterans Hospital
Building 70 (152)
200 Springs Road
Bedford, MA, 01730
United States
Phone: 1 781 687 3181
Fax: 1 781 687 3106
Email: timothy.hogan@va.gov

Abstract

Background: As information and communication technology is becoming more widely implemented across health care organizations, patient-provider email or asynchronous electronic secure messaging has the potential to support patient-centered communication. Within the medical home model of the Veterans Health Administration (VA), secure messaging is envisioned as a means to enhance access and strengthen the relationships between veterans and their health care team members. However, despite previous studies that have examined the content of electronic messages exchanged between patients and health care providers, less research has focused on the socioemotional aspects of the communication enacted through those messages.

Objective: Recognizing the potential of secure messaging to facilitate the goals of patient-centered care, the objectives of this analysis were to not only understand why patients and health care team members exchange secure messages but also to examine the socioemotional tone engendered in these messages.

Methods: We conducted a cross-sectional coding evaluation of a corpus of secure messages exchanged between patients and health care team members over 6 months at 8 VA facilities. We identified patients whose medical records showed secure messaging threads containing at least 2 messages and compiled a random sample of these threads. Drawing on previous literature regarding the analysis of asynchronous, patient-provider electronic communication, we developed a coding scheme comprising a series of a priori patient and health care team member codes. Three team members tested the scheme on a subset of the messages and then independently coded the sample of messaging threads.

Results: Of the 711 messages coded from the 384 messaging threads, 52.5% (373/711) were sent by patients and 47.5% (338/711) by health care team members. Patient and health care team member messages included logistical content (82.6%, 308/373 vs 89.1%, 301/338), were neutral in tone (70.2%, 262/373 vs 82.0%, 277/338), and respectful in nature (25.7%, 96/373 vs 33.4%, 113/338). Secure messages from health care team members sometimes appeared hurried (25.4%, 86/338) but also displayed friendliness or warmth (18.9%, 64/338) and reassurance or encouragement (18.6%, 63/338). Most patient messages involved

either providing or seeking information; however, the majority of health care team member messages involved information provision in response to patient questions.

Conclusions: This evaluation is an important step toward understanding the content and socioemotional tone that is part of the secure messaging exchanges between patients and health care team members. Our findings were encouraging; however, there are opportunities for improvement. As health care organizations seek to supplement traditional encounters with virtual care, they must reexamine their use of secure messaging, including the patient centeredness of the communication, and the potential for more proactive use by health care team members.

(J Med Internet Res 2018;20(3):e82) doi:10.2196/jmir.8801

KEYWORDS

health communication; electronic mail; patient portals; patient-centered care; veterans

Introduction

Background

Patient-Centered Care and Communication

The term "patient centeredness," although still being investigated and refined [1-3], has come to encapsulate the intersection of many priorities and strategies that focus on the unique, individual needs of patients [3]. For example, patient centeredness may simultaneously refer to a broader biopsychosocial perspective on health and illness, a focus on individual patients and the influence that health care provider characteristics or behavior and healing environments can have on the care experience, the sharing of power and responsibility across stakeholders, and the building of therapeutic alliances between patients and providers [4,5]. Perhaps because of the broadness of the concept, health care systems still struggle to translate patient-centered care into practice [1-3,6]. Within the Veterans Health Administration (VA), the Office of Patient-Centered Care and Cultural Transformation (OPCC&CT) has been tasked with leading the system redesign and cultural shift required to provide veterans with care that is more patient centered. Launched in 2012, OPCC&CT defines patient-centered care as care that is "personalized" (tailored to personal goals, history, and lifestyle), "proactive" (preventive care, which leverages holistic approaches), and "patient driven" (led by what matters most to the individual patient) [7,8]. Thus, in VA, patient-centered care is built upon the veteran's experience, such as healing environments of care and genuine, personal relationships with providers, as well as a focus on personalized care across multiple domains of wellness (ie, mind, body, and spirit). Patient-centered communication is an essential component of this care, aiming to strengthen the patient-provider partnership by eliciting and understanding the patients' perspectives, needs, and values; providing patients with the information needed to participate in care to the extent that they desire; and building a shared understanding of a health problem and its treatment [9,10]. In VA, OPCC&CT views patient-provider communication as an essential component in understanding the veteran perspective and in fostering true partnerships between veterans and their providers [8].

A substantial body of research employing a variety of methods [11-16] indicates that patient-centered communication affects various health care processes, patient behaviors, and health outcomes [17]. Patient-centered communication improves processes such as increased patient participation during an encounter [13] and patients' recall of treatment information [18-20]. Influences on more intermediate outcomes are also seen, such as increased satisfaction with care [13,16,21], confidence in communication [22], and improved treatment adherence and appointment follow-up [23]. Moreover, patient-centered communication has been linked to health outcomes such as improved metabolic control and fewer physical limitations in diabetes, better hypertension control, short-term pain control in cancer patients, improved functional status in ulcer patients, less inflammatory organ damage in patients with lupus, and improved emotional well-being [11,17,24]. However, most of this work has focused on the intermittent in-person encounter and has not extended to continuous care supported by technology.

Role of Technology in Patient-Centered Communication

Information and communication technologies have the potential to support patient-centered communication by providing patients with health information to prepare for face-to-face visits and engagement in care [25-27] and strengthening the patient-provider relationship [28,29]. Patient-provider email or asynchronous electronic secure messaging enables patients to interact with their health care providers to exchange nonurgent health information [27,29-31] and has been associated with improved chronic disease self-management [32,33], reduced outpatient visits [34], and urgent care utilization [35]. VA has strategically promoted the use of secure messaging toward the goals of improved communication between veterans and their care teams and increased continuity of care [36]. Nevertheless, the adoption of secure messaging and other asynchronous forms of communication may have inherent disadvantages. Nonverbal modes of communication, which often assist to convey context and tone, are significantly limited by the use of electronic communication such as secure messaging [37,38]. As a result, electronic communication may create more psychological distance between parties, reducing the likelihood of secure messaging being effectively used for interpersonal communication goals such as relationship-building [37-39]. The asynchronous nature of secure messaging may also prevent recipients from receiving immediate feedback or clarity about any interpersonal misunderstandings, which can lead to further miscommunication [37]. Yet, others suggest that users have already adapted their communication styles to adjust to electronic mediums by relying on emoticons or emojis and textual emphasis (eg, ALL CAPS) to project relational content

[37,38]. Similarly, patients have expressed value in the ability to formulate and articulate their questions for providers at their own convenience [28,30], often feeling more comfortable with disclosing personal details due to the psychological distance mentioned earlier [39,40]. Secure messaging has also been found to lower the threshold at which patients initiate communication, resulting in more interactions between physical encounters and perception of greater access [31].

Although previous research has examined the content of electronic communication between patients and providers extensively [30,41-46], less research has focused on the "socioemotional" aspects of the communication enacted by those messages, including how socioemotional tone is expressed and whether it reflects the patient-centered goals of eliciting patients' perspectives, addressing their needs, expressing and responding to emotions, and contributing to a therapeutic partnership [28]. One study of secure messaging identified patient frustration with a perceived lack of empathy in some physician-sent electronic communications [30], suggesting that the socioemotional communication that could facilitate patient-centered goals may be lacking in electronic messages. However, given that the ultimate goal of patient-centered care is to be responsive to patient's needs, what constitutes patient-centered electronic communication may look different for each patient and may need to be tailored to patient-communication styles and preferences. Thus, to further enhance the potential benefits of secure messaging, it is important to explore the various ways that patient centeredness might be realized (or not) through communication via this electronic medium. This is especially important as health care organizations seek to supplement traditional encounters with virtual care.

Research Questions

The VA is a large, geographically diverse, and integrated care system that has a tethered personal health record (PHR) patient portal. VA patients who use secure messaging are also diverse, and penetration of secure messaging into the veteran population is higher than in the general population. As of December 2017, over 2.5 million veteran patients had access to secure messaging, representing a penetration rate of approximately 42% of the 5.9 million VA patients receiving health care services in fiscal year 2017 [47]. Within this context, we conducted a cross-sectional coding evaluation of a corpus of secure messages exchanged between VA patients and health care team members at different VA facilities. Our work contributes to the limited knowledge of how socioemotional tone can be electronically communicated in the secure messages exchanged between patients and health care team members, along with the content and purpose of those messages. Our research questions included (1) why do patients and health care team members exchange secure messages? and (2) what socioemotional tone do these messages convey?

In the Discussion section, we explore how these aspects of the secure messages exchanged between patients and health care team members might facilitate the goals of patient-centered care.

Methods

Study Design and Setting

We identified 8 VA facilities to sample secure messages. Each of these facilities is located in a metropolitan area of the United States; however, because of their catchment areas, they also serve the needs of veterans living in nearby rural areas. We selected these facilities because they are comparable in terms of the diverse patients that they serve, many with complex health care needs, and the wide range of clinical services that they offer.

Patient-Centered Care and Secure Messaging in the Veterans Health Administration

As part of a system-wide transformational initiative [48], VA's Primary Care Program Office has implemented a patient-centered medical home model known as Patient Aligned Care Teams (PACT) [49-51]. The principles of the PACT model call for care that is patient-driven, team-based, efficient, comprehensive, continuous, and encompasses good communication and coordination [52]. The use of information and communication technologies is often considered a critical component of patient-centered medical homes [53]. Within the PACT model, asynchronous secure messaging through the VA's PHR portal, My Health*e*Vet, is envisioned as a means to enhance access to care, support bidirectional communication between patients and health care team members, and supplement other communication mediums [31,54].

In keeping with the PACT model of team-based care, VA implemented secure messaging with a flexible triage team model. Similar to telephone triage, an identified member(s) of the health care team is responsible for reading incoming secure messages and can respond directly or assign action to another member of the triage team, or to another triage team as needed. Health care team members can elect to save all or parts of a secure message or message thread as a progress note in the VA electronic health record (EHR), based on clinical relevance.

Data Collection: Message Corpus

To select secure messaging threads for our evaluation, we used VA administrative datasets to identify all patients at the 8 facilities whose EHR progress notes showed secure messaging threads containing at least 2 messages (an original message and corresponding responses) between January and July 2013. From this sample of patients, we then gathered a random sample of threads, deidentified them, and copied them into a structured template for analysis. In an effort to represent a variety of patient and health care team profiles, our goal was to assemble a sample comprising 50 secure messaging threads from each of the 8 facilities. For the purposes of this evaluation, the secure messaging threads randomly selected for analysis were gathered between September and November 2013. Key demographic characteristics of the patients represented by the sample of secure messaging threads were obtained from VA administrative datasets. The evaluation was exempt from institutional review board review as part of a larger VA quality improvement initiative.

Table 1. Domains for coding secure messages.

Domain	Stakeholder	Codes (subcodes)
Content: The "what" of the message, representing biomedical, holistic, or logistical aspects	Patient and health care team member	Logistical, biomedical, holistic, nonmedical
Socioemotional tone: The "how" of the message, representing the feeling(s) that the message conveys	Patient and health care team member	Attitude (neutral, positive, negative); formality (formal, informal); respectfulness; concern or worry; assertiveness; hurried or rushed; friendliness or warmth; reassurance or encouragement; sympathetic or empathetic; anger or irritation; reflective or legitimizes; depression or sadness
Purpose: The "why" of the message, representing the reason(s) for the message	Patient	Information seeking (proactiveness, treatment or care plan, prescription refill, symptom related, health care team member opinion, test related, referral request, request to fill out form); information provision (health update, responding to health care team member questions); confirmation (gratitude, acknowledgement)
	Health care team member	Information provision (responding to patient questions with pertinent information; giving instructions; providing orientation to medical procedures, therapy, or prevention, checking understanding); information seeking (eliciting patient response regarding treatment or action plan, symptom related, previous treatment plans)

Coding Scheme of Patient-Centered Communication Elements

Drawing on previous literature regarding the analysis of asynchronous, patient-provider electronic communication [28,30,41,42], we developed a series of a priori patient and health care team member codes to apply to the individual messages appearing in each thread. Following our literature review, conversations within our team, and consultations with other patient-provider communication experts, we organized these codes into 3 domains: (1) message content, (2) message socioemotional tone, and (3) message purpose (Table 1). The purpose and tone domains were further categorized to reflect work by Roter and colleagues [4,28], demonstrating that information exchanged between patients and providers carries emotional meaning, cognitive meaning, affective talk, and instrumental behaviors.

Coding Reliability

To assess reliability, 3 team members with expertise in qualitative analysis coded an initial subset of 4 secure messaging threads including 8 individual secure messages. The initial inter-rater reliability (N=3) revealed Cohen kappa values from .80 to 1.00 across all codes [55]. The team members then met to discuss discrepancies and revise the coding scheme with examples to clarify conceptual distinctions and to test the enhanced coding schema with an additional 4 secure messages. The full sample of 384 secure messaging threads was then divided among the 3 team members to code independently using QSR International's NVivo V10 software to support data analysis. The team met weekly to discuss and resolve coding questions. Revisions were made and examples added to the coding scheme as needed. Recognizing that an email message can serve multiple communication functions (eg, information exchange as well as sharing worry or concern) [28], each message in a secure messaging thread could be assigned multiple codes as appropriate, and those codes were not mutually exclusive. In other words, each secure message could be assigned multiple instances of content, socioemotional tone, or purpose. The exceptions to this practice were the codes "formal" and "informal," which were assigned once to each message. Due to the complexity of the coding scheme, intercoder reliability was assessed at weekly meetings before discussing coding questions. The 3 coders maintained their coding reliability of .80 to 1.00 across all codes (average Cohen kappa=.88) throughout the analysis. In the results below, coding frequencies are summarized at the individual message level and the totals can equal more than 100%. Although this does mean that an individual message represents multiple instances of content and socioemotional tone, it follows rigorous coding strategies typical of such analyses [56]. In keeping with contemporary qualitative analysis, our goal was to provide a snapshot of the content and socioemotional tone represented in these secure messages rather than a definitive measurement of prevalence.

Results

Characteristics of Patients and Health Care Team Members

The secure messages in our sample were sent by 292 unique patients and 205 unique VA health care team members across the 8 facilities. As indicated in Table 2, the majority of the patients were male, white, and not of high economic need, meaning that their income was above the threshold set by VA to be eligible for cost-free health care. Their mean age was 59.6 years. VA utilizes a system called the Rural-Urban-Commuting Areas to distinguish between urban and rural areas in the United States based on patient zip codes of residence [57]. Although 86% of the patients sending messages were from urban areas, nearly 14% of the patients were from rural areas. The health care team members responding to the messages were largely registered nurses or physicians; fewer messages were sent by nursing assistants or other team members.

Table 2. Patient and health care team member characteristics.

Patient characteristics[a]	Value
Age (mean, SD)	59.6 (12.3)
Gender, n (%)	
Male	242 (85.5)
Female	41 (14.5)
Race, n (%)	
White	207 (73.1)
African-American	31 (11.0)
Unknown or missing	39 (13.8)
Other	6 (2.1)
Socioeconomic status, n (%)	
High economic need	62 (21.9)
Geographic location, n (%)	
Urban	245 (86.6)
Rural	38 (13.4)
Elixhauser comorbidity index (mean, SD)	2.9 (2.3)
Role of health care team member, n (%)	
Registered nurse	84 (41.0)
Physician	65 (31.7)
Nursing assistant	30 (14.6)
Other	10 (4.9)
Advanced practice nurse	9 (4.4)
Physician assistant	3 (1.5)
Psychologist	2 (1.0)
Medical assistant	1 (<1.0)
Social worker	1 (<1.0)

[a]Data missing for 9 patients.

Message Characteristics

Across the 8 facilities, there were differences in the number of secure messaging threads that health care team members had elected to save into the EHR as a progress note over the selected 6-month evaluation period. As such, we were able to gather an average of 48 secure messaging threads from each facility (min=37, max=51) for a total sample of 384 threads comprising 711 individual secure messages. Of the 384 secure messaging threads, most were initiated by patients (90.9%, 349/384) rather than a health care team member (9.1%, 35/384). Of the 711 individual messages, roughly half were sent by patients (52.5%, 373/711), and half were sent by health care team members (47.5%, 338/711). Finally, most patient messages appeared to be composed by the patient him/herself (92.2%, 344/373) as opposed to a proxy (eg, a family member or other informal caregiver; 7.8%, 29/373).

Message Content

Table 3 presents our content codes with exemplary quotes from patient and health care team member messages.

Messages Sent by Patients

The majority of patient secure messages included logistical content (82.6%, 308/373) such as scheduling an appointment or requesting a prescription refill. Half of the patient messages also included biomedical content (50.4%, 188/373), such as mentioning specific diseases, medications, or treatments. Almost 10% of the patient messages (8.6%, 32/373) included holistic content, discussing psychosocial aspects of health, such as exercise, stress management, or family relationships. Fewer messages contained nonmedical content (6.4%, 24/373) such as mention of a change of email address.

Messages Sent by Health Care Team Members

The majority of health care team member secure messages were largely logistical in nature (89.1%, 301/338), and many contained biomedical content (29.6%, 100/338). Less than 5% of the messages sent by health care team members were coded as containing holistic content (4.4%, 15/338) but 8.0% contained nonmedical content (27/338), such as acknowledging a holiday or an occasion in one's personal life.

Table 3. Examples of patient and health care team member message content.

Code and stakeholder	Sample message excerpt	Code presence, n (%)
Logistical content		
Patient	"I will be at [location] on Tuesday, March 26 and would like to go to the Dental Clinic to begin some long overdue dental work"	308 (82.6)
Health care team member	"Wheelchair referral has been placed. Dental consult only good for 72 hours - so I can place the consult closer to the time you would like to go to the clinic"	301 (89.1)
Biomedical content		
Patient	"When you're back on duty, please be good enough to enter a refill for me for Lisinopril 40MG. My BP lately ranges from 120's to low 150's over 60's / 70's"	188 (50.4)
Health care team member	"Upon review of your records both x-rays and medications; your x-rays show that you have early degenerative changes to both knees (arthritis)"	100 (29.6)
Nonmedical content		
Patient	"Here is that link I promised for the 'Battlefield Of The Mind' documentary. Fortunately, I am not in it but I contributed money to its making. If you like it after watching, Please buy a DVD copy at the second website. It's only a few dollars."	24 (6.4)
Health care team member	"Hi [name], Happy Mother's Day to you as well! I hope you had a great one."	27 (8.0)
Holistic health content		
Patient	"I don't know what else to do. I don't want to quit work without knowing that financially I can't support my family. I don't want to cause any more stress on myself although work in itself is stressful"	32 (8.6)
Health care team member	"I have entered a consult for the Move program. This is a weight management program for veterans"	15 (4.4)

Message Socioemotional Tone

Tables 4 and 5 presents our socioemotional tone codes with exemplary quotes from patient and health care team member messages.

Messages Sent by Patients

Patient messages were frequently coded as neutral in tone (70.2%, 262/373), being direct, to-the-point, and transactional. The remainder of the patient messages were equally positive (14.2%, 53/373) or negative (14.2%, 53/373) in tone. Positive messages reflected patient optimism toward health conditions or treatment plans. Negative messages reflected pessimism stemming from perceptions of health conditions, treatment plans, or actual treatments received. In terms of emotions, almost half of the messages expressed some concern or worry (39.9%, 149/373). Patient messages also exhibited respectfulness (25.7%, 96/373), being mannerly and considerate of the potential feelings and situations of health care team members. However, messages were also equally assertive (25.5%, 95/373) or direct. Many messages conveyed friendliness or warmth (14.5%, 54/373), reading as chatty or chummy and attempting to engage health care team members. Fewer messages demonstrated being reassured or encouraged (1.9%, 7/373), where patients expressed relief and an optimistic outlook about their health or treatments. Patient messages were also coded as more informal (61.1%, 228/373), lacking proper grammar or a salutation and signature, rather than formal (38.9%, 145/373).

Messages Sent by Health Care Team Members

Similar to patient messages, the tone of health care team member messages was largely neutral (82.0%, 277/338) rather than positive (14.2%, 48/338) or negative (2.7%, 9/338), and exhibited respectfulness (33.4%, 113/338). At times, the messages appeared hurried or rushed (25.4%, 86/338), although a substantial portion displayed friendliness or warmth (18.9%, 64/338) and offered reassurance or encouragement to patients (18.6%, 63/338). More than half of the secure messages sent by health care team members were coded as informal (59.2%, 200/338) rather than formal (40.8%, 138/338).

Message Purpose

Tables 6 and 7 present our purpose codes with exemplary quotes from patient and health care team member messages.

Messages Sent by Patients

Most patient messages involved either providing or seeking information. In terms of information provision, just under half of the messages were coded as a health update (48.8%, 182/373) in which the patient informed a health care team member about some aspect of their current health and well-being.

Regarding information seeking, patient messages requested information regarding treatment or care plans (22.5%, 84/373), prescription refills (22.0%, 82/373), symptoms (16.1%, 60/373), or test results (13.7%, 51/373). Almost a quarter of the messages were coded as proactive in nature (23.9%, 89/373), where patients took initiative to ask questions, express disagreement, or actively contribute to the management of their care. Patients also sought the opinions of health care team members regarding various care-related issues (15.5%, 58/373). Distinct from providing and seeking information, a smaller number of patients sent confirmatory messages to acknowledge receipt of a health care team member message (1.6%, 6/373) or to express gratitude to them (7.0%, 26/373).

Messages Sent by Health Care Team Members

The majority of the secure messages sent by health care team members responded to patient questions with information of some kind (72.8%, 246/338). Health care team members also used secure messaging for giving instructions or providing patients with specific action steps toward care (30.5%, 103/338). Over a quarter of the messages also offered orientation to procedures, therapies, or prevention behaviors (26.3%, 89/338). Few of the health care team member messages reflected information seeking; only 5.6% (19/338) were coded as eliciting a patient response regarding plans for treatment or a future course of action. In addition, few health care team members utilized secure messaging as a way to ask patients about symptoms (3.3%, 11/338) or previous treatment plans (3.0%, 10/338).

Table 4. Examples of patient message tone.

Code	Sample message excerpt	Code presence, n (%)
Neutral attitude	"Just wanted you to know that the MRI is scheduled for June 4, 2013 at 10:30 AM"	262 (70.2)
Positive attitude	"Good news! I'm up and getting around some with the ortho boot and walker. In fact, yesterday and today I actually made it outside over a high doorsill and one step on my own! 3 times today!"	53 (14.2)
Negative attitude	"We tried that already and it didn't work. I realize that things have to be shown not to work before they are changed, but in the mean time I am still gagging and getting headaches"	53 (14.2)
Informal	"I need another holder for the eye drops as the rubber seam still splits around the bottle. Can you pls reorder?"	228 (61.1)
Formal	"[Dr. name], I request a refill of my monthly supply of: (RX# [prescription number]) ACETAMIN 325MG/OXYCODONE 5MG TAB, dispensed on 5 April 13. I am available to pick up medication at the [location] Clinic Pharmacy on 3 May 13 due to the 5th of May is a Sunday. Thank you, [patient name]"	145 (38.9)
Concern or worry (includes anxiety or nervousness)	"[Dr. name], I have been anxiously waiting on your call since this morning. I called the Heart Clinic and the Echocardiogram has been read, dictated and is in the system. PLEASE call me with these results. I've been sick with worry."	149 (39.9)
Respectfulness	"[Dr. name], I am flying out of town, for work, Monday at 1:00pm. If possible, I would like pick my monthly Methadone prescription at the Pharmacy window, Monday morning at 9:00am? As always, thank you for your help. Respectfully, [Patient name]"	96 (25.7)
Assertiveness	"I need you to put in my order for the lab to take blood. I thought [Dr.name] had done it, but there's no order."	95 (25.5)
Friendliness or warmth	"Good afternoon [Dr. name], hope all is well. I am requesting to have the following medications renewed: Diclofenac and Pravastatin. Also, I received my card for my 6-month follow-up, so if possible, I would like to set up that appointment at your earliest convenience. Thank you, have a good day!"	54 (14.5)
Hurried or rushed	"checking on status of morphine rx.. also need Dilantin, zomig, and ceterzine refilled, thx."	33 (8.8)
Anger or irritation	"I do not need gauze sponges and it seems every time I try to get drain sponges I get gauze sponges. I at least need to have the order for drain sponges available so I can go to pharmacy and pick them up to avoid further problems. These items are very similar and this happens all the time because they are so similar. But I need DRAIN SPONGES."	26 (7.0)
Depression or sadness	"My Dad passed away on the 9th and I am having a hard time. I feel so empty and lost. I miss him so much."	9 (2.4)
Reassurance or encouragement	"The extra dose of Lopressor seems to be working. Thanks for the new BP machine."	7 (1.9)
Sympathetic or empathetic	—	—
Reflective or legitimizes	—	—

Table 5. Examples of health care team member message tone.

Code	Sample message excerpt	Code presence, n (%)
Neutral attitude	"Your medication has been refilled"	277 (82.0)
Positive attitude	"I'm glad that your range of motion is improving, even if only slightly to begin with"	48 (14.2)
Negative attitude	"I referred you to rehab. I have no quick answer for your pain. With your chronic osteoarthritis of the knees, knee pain will always be there. The goal is to bring the pain level down so you can function better but to get rid of it totally, this may not be a realistic goal"	9 (2.7)
Informal	"done, and given to pharmacy"	200 (59.2)
Formal	"[Patient name], I hope you are well. I see that you did not make your appointment to the endocrinologist. I believe we need to get their opinion as well and then I would like to see you again. Most sincerely, [Dr. name]"	138 (40.8)
Respectfulness	"Good morning, I will renew both the pseudoephedrine and saline for mail. Have a great week, [Dr. name]"	113 (33.4)
Hurried or rushed	"I have written the scripts and will be sent to VA pharmacy today."	86 (25.4)
Friendliness or warmth	"I will have our clerk get you scheduled. They may be in the process. That is a week of vacation for me that accidentally wasn't blocked earlier this year. Sorry about rescheduling. We will send you a new appt. Hope your wife has a full, speedy recovery."	64 (18.9)
Reassurance or encouragement	"Feel free to message me whenever you need to – you are not bugging me!"	63 (18.6)
Assertiveness	"Your Lantus Rx has no more refills and has to be renewed. You should have enough to cover you till close to end of March according to your chart. [Dr name] will be made aware in order to renew and have it mailed to you. Thank you."	41 (12.1)
Sympathetic or empathetic	"I am so sorry to hear you have not been feeling well. I will give you a call to discuss."	28 (8.3)
Reflective or legitimizes	"Hi [son's name], I do think it is reasonable to consider rivastigmine, but the VA does not yet have the transdermal patch. We do have the pill formulation which I believe has a slightly higher rate of side effects."	17 (5.0)
Concern or worry (includes anxiety or nervousness)	"I am sorry to hear that you fell- please come to the ER if it happens again. Did you end up going to a local hospital?"	15 (4.4)
Anger or irritation	"I am not exactly sure why u are emailing me every day about your nutritional data. I do not know who asked you to do this. I know that I have not."	3 (0.9)
Depression or sadness	—	—

Table 6. Examples of patient message purpose.

Code and subcode	Sample message excerpt	Code presence, n (%)
Information-seeking		
Proactiveness	"[Dr name], I have used up all of the Clotrimazole you prescribed for me. I still have the itching on the middle portion of my body. Is there something else that will work better, a spray or something like that?"	89 (23.9)
Treatment or care plan	"Hello, [Dr. name], When I was at your office yesterday my blood pressure was high. I checked it today and it's still running high: 147/95. Maybe it's time for a new blood pressure medication. I have gotten older since you prescribed Lisinopril 5mg."	84 (22.5)
Prescription refill	"I also need a prescription for my nitroglycerine tablets. My current supply is about to expire."	82 (22.0)
Symptom related	"I would like to make an appointment to check on a swelling that is taking place below and to the right of my tongue. No pain or sensations, just an obvious swelling beneath the outside skin."	60 (16.1)
Health care team member opinion	"They also want me to change the Meloxicam for Tramadol, Naproxen and time-scheduled Tylenol. This was after I wrote and requested the 90 prescription of Meloxicam. I would truly like to hear your opinion on this."	58 (15.5)
Test related	"By any chance, have my HIV results come back yet?"	51 (13.7)
Referral request	"I was wondering if you can put in referrals for me for Neurology and Endocrinology. I need to see someone about the migraines and also about my pituitary growth."	22 (5.9)
Request to fill out form	"My job gave me the form for disability and there is a portion for you to complete"	18 (4.8)
Information provision		
Health update	"Another interesting factoid: yesterday, I weighed myself. I got out of the shower and I weighed 170 pounds! So that means, since May, I lost about 20 to 25 pounds."	182 (48.8)
Responding to health care team member questions	"It was many years ago, maybe 3 or 4. Don't remember the dosage."	9 (2.4)
Confirmation		
Gratitude	"Both my wife and I really want to thank you for your patience and care last week. Although you may feel that you were just doing your job, to us, it meant so much."	26 (7.0)
Acknowledgment	"Thank you. I will make this work."	6 (1.6)

Table 7. Examples of health care team member message purpose.

Code and subcode	Sample message excerpt	Code presence, n (%)
Information provision		
Responding to patient questions with pertinent information	"Just got it back. It is normal. You should recheck in 4 months."	246 (72.8)
Giving instructions	"Please stop by 6C to give a urine sample and also to have [name] or any Medical Assistant check your blood pressure and record - it was a bit high on recent check."	103 (30.5)
Providing orientation to medical procedures, therapy, or prevention	"Although the pulses in your feet are fine, we can send you to the [location] VA for Ankle Brachial Indices testing. This tests your blood pressure in your upper extremities relative to your lower extremities (down to your toes). This is a first-line test in evaluating the circulation in your legs."	89 (26.3)
Checking understanding	"I thought we were going to do it through the hematology clinic given your previous events? Adding [Dr. name] for opinion."	20 (5.9)
Information seeking		
Eliciting patient response regarding treatment or action plan	"You have arthritis in the knees - would you like a referral for exercise therapy?"	19 (5.6)
Symptom related	"Are you having any vision problems now?"	11 (3.3)
Previous treatment plans	"Januvia, or sitagliptin, is a restricted drug. I can place a nonformulary request if you like. If so, I need to know what diabetes meds you have tried that did not work out."	10 (3.0)

Discussion

Principal Findings

We analyzed a sample of secure messages between VA patients and health care team members to understand why these messages are exchanged, and what socioemotional tone the messages convey. Overall, our findings regarding message content are consistent with prior studies. However, our findings related to message tone and implications for emotional expression extend the existing literature.

Message Content

We examined message content to place our findings about the tone and purpose of the messages in our sample into a broader context. Our analysis revealed that the majority of patient secure messages included logistical content, demonstrating the organizing efforts of patients to ensure that they had the resources (eg, prescriptions) needed for their own care. These findings are consistent with previous studies, including some studies conducted in VA, highlighting the frequent use of email and secure messaging by patients to address administrative issues and related care actions [29,41,42,44-46]. Although there may be a tendency to view such content as uncomplicated or routine, we believe that it underscores the importance of secure messaging as a tool that patients use to promote care coordination. Patients must often play an active role in coordination [58]; the prevalence of logistical content in our message sample indicates that electronic communication is commonly used by patients to facilitate at least some of this work.

Half of the patient messages also included specific biomedical content, most often the formal names of health conditions and prescription medications, and considerably less holistic and nonmedical content. Health care team member messages were similar, not surprisingly, given that most were responses to secure messages initiated by patients. This point raises important questions about the intended uses of secure messaging, an issue we examine below.

Message Socioemotional Tone

Few prior studies have evaluated message tone, a notable exception being Roter et al [28], who found that in a convenience sample of email exchanges between 8 doctor-patient dyads, patients often used email to convey their emotional state, frequently an expression of worry or concern. Our analysis extends beyond Roter et al and found that both the patient and the health care team member secure messages were largely neutral in tone and also tended to exhibit an informal style. This is not to say, however, that messages were devoid of emotion. On the contrary, many messages initiated by patients expressed concern and worry, warmth, or even anger. An important lesson from our analysis is that the neutral tone and informal style characterizing some patient messages should not distract health care providers from the emotions that patients may convey through this medium.

Several positive tonal elements were prominent in the health care team member messages in our sample, including being respectful of patients, showing friendliness or warmth toward them, and offering them reassurance or encouragement. These findings again map to Roter et al [28], who found that many physicians expressed concern, reassurance, partnership, and other supportive socioemotional expressions in response to patient messages containing emotional content. Socioemotional tone can function in interpersonal communication to build rapport and strengthen the relationship between parties [28,59], contributing to the therapeutic partnership that is central to patient-centered care. Still, one-fourth of the health care team member messages in our sample were coded as hurried or rushed. Taken together, these findings suggest that there is considerable variation in the tone of secure messages sent by

VA health care team members. It is important to frame our findings about the tone of health care team member messages in the context of the workflow that surrounds secure messaging in VA. Per VA policy, team members are expected to respond to patient messages within 3 business days, and triaging approaches are often used to assign messages to appropriate team members to ensure efficient responses. These realities, coupled with large patient panels and tightly scheduled clinics, may explain the number of messages coded as informal, hurried, or rushed, and those that appeared less sympathetic, reflective, or legitimizing of patient concerns.

Message Purpose

Nearly half of the patient messages in our sample were coded as information updates, in which they informed health care team members about some aspect of their health. Patient provision of such updates has been documented in previous studies [42,44,46] and is a powerful illustration of how secure messaging can facilitate the shift from episodic to continuous care and cultivate ongoing, healing relationships as argued for by the Institute of Medicine [60]. In addition to providing updates, patients also used secure messaging to seek information about a variety of topics. Although some of these topics, such as prescription refills and test results, reflect the high prevalence of logistical content in our messages, other topics, such as symptoms and team member opinions, suggest that patients also use secure messaging to seek information about more nuanced topics.

Our exploration of message purpose did reveal gaps. The numerous instances of information provision coded in health care team member messages is not surprising—the majority involved responding to patient questions with information, and over a quarter gave instructions or offered orientation to some health-related topic. Eliciting the patient's perspective has been described as an important element of patient-centered communication [9]; however, there were few instances in our data where health care team members appeared to use secure messaging to reach out to patients and seek information from them regarding treatment plans. Even in their replies, it was uncommon for health care team members to ask for patient input on topics such as treatment plans or descriptions of symptoms. These trends highlight the reactive nature that tends to characterize much secure messaging use among health care team members, similar to in-person encounters.

Practice Implications

Our analysis offers a snapshot of the electronic communication between the patients and health care team members represented by the secure messages in our sample. Although the literature has suggested that being responsive to patient emotions and concerns can build rapport and contribute to a therapeutic relationship [28,59], we must caution that our evaluation is not intended to serve as a determination of whether the secure-messages in our sample were in fact "patient-centered." As discussed above, we identified considerable variation in tone among health care team member messages. However, placing our findings in the context of patient-centered communication may suggest ways that the content and socioemotional tone of these secure messages could facilitate the goals of patient-centered care. For example, we coded ample instances of health care team members being respectful, sympathetic, friendly, reflective, and reassuring, which could indicate a response to patient-expressed emotion. Yet, we also coded many instances of health care team members seeming hurried or rushed in their messages. Although at face value, messages coded as the former could be considered "more patient-centered" and messages coded as the latter could be considered "less patient-centered," we must remember that patient centeredness can encompass a wide range of behaviors. It is likely that some patients would prefer receiving messages from their health care team members that reflect elements of friendliness or sympathy; however, there are likely other patients who would prefer receiving messages that are succinct, timely, and to-the-point; messages that, in our sample, would likely have been coded as hurried or rushed. These points underscore the importance of understanding and embracing patient preferences for communicating with members of their health care team and using technologies such as secure messaging. It may be that true patient centeredness in secure messaging involves health care team members discussing upfront with their patients what exactly they want from their communication in this medium and, in turn, tailoring their approach to that individual patient. Future work should examine ways of eliciting patient preferences for secure messaging, testing different approaches for implementing those preferences into practice, and assessing their impact on patient satisfaction and other outcomes.

Although what might be considered "patient-centered" content and socioemotional tone in secure messaging is likely to vary with patient preferences, we suggest that there are ways in which elements of patient-centered communication can be further integrated into asynchronous electronic communication to promote the therapeutic relationship. These opportunities are critical given the goals of many health care systems to increase secure messaging use and to expand virtual care. The variation in tone of health care team member messages points to the importance of cultivating secure messaging practices that fully elicit the patient's perspective, and empower the patient to participate in their care to the extent they desire. Doing so requires recognition that patients and health care team members are both active producers of meaning when using secure messaging [61], and that messages are themselves more than simple chunks of information. On the contrary, in addition to their purpose, many of the patient messages in our sample included expressions of socioemotional tone. The fact that patient secure messages could at once be conveying emotions, offering important contextual detail, and attempting to accomplish care-related tasks underscores the complexity that can characterize communication in this medium. A multiple-goals perspective [62] that recognizes that patients and health care team members can have a variety of goals when they send a secure message may be a valuable way to frame future secure messaging studies and to understand its use in practice. Applying such a perspective would enable a richer understanding of the complexity of secure messaging communication while also offering analytical tools (eg, types of goals, types of responses to goals) to support more nuanced analyses.

Similarly, the preponderance of secure messages in our sample in which health care team members responded to a patient request or inquiry, and the limited use of the medium by health care team members to reach out to patients or to seek information from them represents a significant missed opportunity to promote patient participation, engagement, and relationship building. Advocating for more "proactive" uses of secure messaging in which health care team members initiate communication, elicit patients' perspectives, and draw them into relevant dialogue would constitute a paradigm shift in current approaches to this communication medium. Health care team members and patients alike will need different training about secure messaging if it is envisioned as much as a medium for engaging patients and bolstering the therapeutic relationship as a medium for addressing logistical needs [28,29]. Corresponding workflow implications for the health care team will also have to be examined.

Limitations

There are several limitations to our evaluation. As noted earlier, the messages that were included in our analysis were only those that the VA health care team members determined clinically relevant to save in the EHR. Additionally, although our goal was to assemble a sample comprising 50 secure messaging threads from each of the participating facilities, in 1 facility, there were fewer than 50 threads in the EHR for the evaluation's selected time period. Our evaluation also focused solely on secure messaging to the exclusion of other communication mediums. We are unable to ascertain whether other types of communication transpired before, during, or after a secure messaging exchange. As others have similarly argued [46], a different analysis that situates secure messaging in the context of other communication mediums and focuses on how those mediums could augment one another would provide additional insights about the content and socioemotional tone evident in secure messages and how secure messaging is being used to address particular patient needs. Finally, although we report basic information about the veteran patients and health care team members who are represented by the secure messages in our sample, we did not construct the sample to explore associations between the patient and health care team member characteristics and our coding domains.

Acknowledging these limitations, we believe there is value in offering the following best practices to inform health care team members' use of asynchronous, electronic secure messaging with patients. These best practices extend previous guidelines for electronic communications and the use of email with patients [63,64]:

- Elicit and understand the preferences each patient may have for communicating with their health care team members through secure messaging. As noted above, what constitutes patient centeredness for one patient may be different for another. Health care team members should discuss the use of secure messaging with each of their patients, preferably at the time the patient is adopting the technology, to set expectations and discuss what the patient hopes to ascertain from communication through this medium. In addition to addressing the content and tone that is part of secure message exchanges, such discussions can also foster patient understanding of a health care system's approach to triaging and processing messages.
- Recognize that expressions of emotion can be an inherent part of patient secure messages. Patients may have various goals in mind when they send a secure message to members of their health care team, and the expression of emotion may be part of those goals. Health care team members should appreciate the presence of emotion as one of the complexities of communicating with their patients through this medium.
- Utilize, as appropriate, patient expressions of emotion as a means to enhance the therapeutic relationship. If the expression of emotion is a natural part of some patient secure messages, not addressing those emotions could be a missed opportunity for health care team members to engage with patients and to use them as a means to foster rapport, shared understandings, and engagement.
- Leverage asynchronous, electronic secure messaging as a means to reach and engage patients. In the current paradigm, much secure messaging use is reactive in nature. A more proactive approach that involves health care team members initiating communication with patients through this medium to seek information and elicit their perspectives could be an effective means of fostering participation in the care process to the extent the patient desires.

Conclusions

Our evaluation represents an important step toward understanding the content and socioemotional tone that is part of the secure messages exchanged between patients and health care team members, and how asynchronous communication might facilitate the goals of patient-centered care. Our findings suggest that there are opportunities to enhance communication in this medium. The rapid implementation of secure messaging across health care systems places a premium on pursuing such improvements in the short run so that desirable process outcomes and longer term clinical outcomes can be realized through its use.

Acknowledgments

This work was funded by the United States Veterans Health Administration, Office of Patient-Centered Care and Cultural Transformation, and Quality Enhancement Research Initiative Program, Grant #PCE 13-001.

We would like to thank Laura Krejci and Tracy Gaudet from the Office of Patient-Centered Care and Cultural Transformation for their support of this work, and Juliet Wu for her assistance in preparing the manuscript for submission.

The contents of this manuscript do not necessarily represent the views of the Department of Veterans Affairs or the US government.

Conflicts of Interest

None declared.

References

1. Berwick DM. What 'patient-centered' should mean: confessions of an extremist. Health Aff (Millwood) 2009;28(4):w555-w565. [doi: 10.1377/hlthaff.28.4.w555] [Medline: 19454528]
2. Epstein RM, Fiscella K, Lesser CS, Stange KC. Why the nation needs a policy push on patient-centered health care. Health Aff (Millwood) 2010 Aug;29(8):1489-1495 [FREE Full text] [doi: 10.1377/hlthaff.2009.0888] [Medline: 20679652]
3. Epstein RM, Street RL. The values and value of patient-centered care. Ann Fam Med 2011;9(2):100-103 [FREE Full text] [doi: 10.1370/afm.1239] [Medline: 21403134]
4. Roter DL, Hall JA, Katz NR. Relations between physicians' behaviors and analogue patients' satisfaction, recall, and impressions. Med Care 1987 May;25(5):437-451. [Medline: 3695654]
5. Mead N, Bower P. Patient-centredness: a conceptual framework and review of the empirical literature. Soc Sci Med 2000 Oct;51(7):1087-1110. [Medline: 11005395]
6. Bardes CL. Defining "patient-centered medicine". N Engl J Med 2012 Mar 1;366(9):782-783. [doi: 10.1056/NEJMp1200070] [Medline: 22375968]
7. US Department of Veterans Affairs. 2016. Office of Patient Centered Care and Cultural Transformation: Promoting Whole Health URL: https://www.va.gov/PATIENTCENTEREDCARE/about.asp [accessed 2018-01-30] [WebCite Cache ID 6wrmn1ehA]
8. US Department of Veterans Affairs. 2015. Capturing Personalized, Proactive, Patient-Driven Care Across VA: A Report by the VA Office of Patient Centered Care and Cultural Transformation URL: https://www.va.gov/patientcenteredcare/FY15AnnualNarrative/StaffFacing/OPC_Staff_Narrative_FY15-508.pdf [accessed 2018-01-30] [WebCite Cache ID 6wrpqKewl]
9. Epstein RM, Franks P, Fiscella K, Shields CG, Meldrum SC, Kravitz RL, et al. Measuring patient-centered communication in patient-physician consultations: theoretical and practical issues. Soc Sci Med 2005;61(7):1516-1528. [Medline: 16005784]
10. Levinson W, Lesser CS, Epstein RM. Developing physician communication skills for patient-centered care. Health Aff (Millwood) 2010 Jul;29(7):1310-1318. [doi: 10.1377/hlthaff.2009.0450] [Medline: 20606179]
11. Cegala DJ. Patient communication skills training: a review with implications for cancer patients. Patient Educ Couns 2003;50(1):91-94. [Medline: 12767592]
12. Mead N, Bower P. Patient-centred consultations and outcomes in primary care: a review of the literature. Patient Educ Couns 2002 Sep;48(1):51-61. [Medline: 12220750]
13. Harrington J, Noble LM, Newman SP. Improving patients' communication with doctors: a systematic review of intervention studies. Patient Educ Couns 2004 Jan;52(1):7-16. [Medline: 14729285]
14. Swenson SL, Buell S, Zettler P, White M, Ruston DC, Lo B. Patient-centered communication: do patients really prefer it? J Gen Intern Med 2004 Nov;19(11):1069-1079 [FREE Full text] [doi: 10.1111/j.1525-1497.2004.30384.x] [Medline: 15566435]
15. Berkhof M, van Rijssen HJ, Schellart AJ, Anema JR, van der Beek AJ. Effective training strategies for teaching communication skills to physicians: an overview of systematic reviews. Patient Educ Couns 2011 Aug;84(2):152-162. [doi: 10.1016/j.pec.2010.06.010] [Medline: 20673620]
16. Henry SG, Fuhrel-Forbis A, Rogers MA, Eggly S. Association between nonverbal communication during clinical interactions and outcomes: a systematic review and meta-analysis. Patient Educ Couns 2012 Mar;86(3):297-315. [doi: 10.1016/j.pec.2011.07.006] [Medline: 21824738]
17. Street RL. How clinician-patient communication contributes to health improvement: modeling pathways from talk to outcome. Patient Educ Couns 2013 Sep;92(3):286-291. [doi: 10.1016/j.pec.2013.05.004] [Medline: 23746769]
18. Robinson EJ, Whitfield MJ. Improving the efficiency of patients' comprehension monitoring: a way of increasing patients' participation in general practice consultations. Soc Sci Med 1985;21(8):915-919. [Medline: 4071125]
19. Lewis CC, Pantell RH, Sharp L. Increasing patient knowledge, satisfaction, and involvement: randomized trial of a communication intervention. Pediatrics 1991 Aug;88(2):351-358. [Medline: 1861939]
20. McGee DS, Cegala DJ. Patient communication skills raining for improved communication competence in the primary care medical consultation. J Appl Commun Res 1998 Nov;26(4):412-430. [doi: 10.1080/00909889809365517]
21. Ong LM, Visser MR, Lammes FB, de Haes JC. Doctor-patient communication and cancer patients' quality of life and satisfaction. Patient Educ Couns 2000 Sep;41(2):145-156. [Medline: 12024540]
22. Tran AN, Haidet P, Street Jr RL, O'Malley KJ, Martin F, Ashton CM. Empowering communication: a community-based intervention for patients. Patient Educ Couns 2004 Jan;52(1):113-121. [Medline: 14729298]
23. Cegala DJ, Marinelli T, Post D. The effects of patient communication skills training on compliance. Arch Fam Med 2000 Jan;9(1):57-64. [Medline: 10664643]

24. Vogel BA, Leonhart R, Helmes AW. Communication matters: the impact of communication and participation in decision making on breast cancer patients' depression and quality of life. Patient Educ Couns 2009 Dec;77(3):391-397. [doi: 10.1016/j.pec.2009.09.005] [Medline: 19796910]
25. Cimino JJ, Patel VL, Kushniruk AW. The patient clinical information system (PatCIS): technical solutions for and experience with giving patients access to their electronic medical records. Int J Med Inform 2002 Dec 18;68(1-3):113-127. [Medline: 12467796]
26. Fisher B, Bhavnani V, Winfield M. How patients use access to their full health records: a qualitative study of patients in general practice. J R Soc Med 2009 Dec;102(12):539-544 [FREE Full text] [doi: 10.1258/jrsm.2009.090328] [Medline: 19966130]
27. White A, Danis M. Enhancing patient-centered communication and collaboration by using the electronic health record in the examination room. J Am Med Assoc 2013 Jun 12;309(22):2327-2328 [FREE Full text] [doi: 10.1001/jama.2013.6030] [Medline: 23757080]
28. Roter DL, Larson S, Sands DZ, Ford DE, Houston T. Can e-mail messages between patients and physicians be patient-centered? Health Commun 2008;23(1):80-86. [doi: 10.1080/10410230701807295] [Medline: 18443995]
29. Ye J, Rust G, Fry-Johnson Y, Strothers H. E-mail in patient-provider communication: a systematic review. Patient Educ Couns 2010 Aug;80(2):266-273 [FREE Full text] [doi: 10.1016/j.pec.2009.09.038] [Medline: 19914022]
30. Houston TK, Sands DZ, Jenckes MW, Ford DE. Experiences of patients who were early adopters of electronic communication with their physician: satisfaction, benefits, and concerns. Am J Manag Care 2004 Sep;10(9):601-608 [FREE Full text] [Medline: 15515992]
31. Nazi KM. The personal health record paradox: health care professionals' perspectives and the information ecology of personal health record systems in organizational and clinical settings. J Med Internet Res 2013;15(4):e70 [FREE Full text] [doi: 10.2196/jmir.2443] [Medline: 23557596]
32. Ralston JD, Revere D, Robins LS, Goldberg HI. Patients' experience with a diabetes support programme based on an interactive electronic medical record: qualitative study. Br Med J 2004 May 15;328(7449):1159 [FREE Full text] [doi: 10.1136/bmj.328.7449.1159] [Medline: 15142919]
33. McMahon GT, Gomes HE, Hickson Hohne S, Hu TM, Levine BA, Conlin PR. Web-based care management in patients with poorly controlled diabetes. Diabetes Care 2005 Jul;28(7):1624-1629 [FREE Full text] [Medline: 15983311]
34. Zhou YY, Garrido T, Chin HL, Wiesenthal AM, Liang LL. Patient access to an electronic health record with secure messaging: impact on primary care utilization. Am J Manag Care 2007 Jul;13(7):418-424 [FREE Full text] [Medline: 17620037]
35. Shimada SL, Hogan TP, Rao SR, Allison JJ, Quill AL, Feng H, et al. Patient-provider secure messaging in VA: variations in adoption and association with urgent care utilization. Med Care 2013 Mar;51(3 Suppl 1):S21-S28. [doi: 10.1097/MLR.0b013e3182780917] [Medline: 23407007]
36. US Department of Veterans Affairs. 2018. FAQs - MyHealtheVet URL: https://www.myhealth.va.gov/mhv-portal-web/web/myhealthevet/faqs [accessed 2018-01-30] [WebCite Cache ID 6wrqoSrLx]
37. Byron K. Carrying too heavy a load? The communication and miscommunication of emotion by email. Acad Manage Rev 2008;33(2):309-327. [doi: 10.5465/?AMR.2008.31193163]
38. Thurlow C, Lengel LM, Tomic A. Computer Mediated Communication: Social Interaction and the Internet. London: SAGE Publications Ltd; 2004.
39. Baur C. Limiting factors on the transformative powers of e-mail in patient-physician relationships: a critical analysis. Health Commun 2000;12(3):239-259. [doi: 10.1207/S15327027HC1203_02] [Medline: 10938915]
40. Strasser F, Fisch M, Bodurka DC, Sivesind D, Bruera E. E-motions: email for written emotional expression. J Clin Oncol 2002 Aug 01;20(15):3352-3355. [doi: 10.1200/JCO.2002.20.15.3352] [Medline: 12149311]
41. Sittig DF. Results of a content analysis of electronic messages (email) sent between patients and their physicians. BMC Med Inform Decis Mak 2003 Oct 01;3:11 [FREE Full text] [doi: 10.1186/1472-6947-3-11] [Medline: 14519206]
42. White CB, Moyer CA, Stern DT, Katz SJ. A content analysis of e-mail communication between patients and their providers: patients get the message. J Am Med Inform Assoc 2004;11(4):260-267 [FREE Full text] [doi: 10.1197/jamia.M1445] [Medline: 15064295]
43. North F, Crane SJ, Stroebel RJ, Cha SS, Edell ES, Tulledge-Scheitel SM. Patient-generated secure messages and eVisits on a patient portal: are patients at risk? J Am Med Inform Assoc 2013;20(6):1143-1149 [FREE Full text] [doi: 10.1136/amiajnl-2012-001208] [Medline: 23703826]
44. Mirsky JB, Tieu L, Lyles C, Sarkar U. A mixed-methods study of patient-provider e-maill content in a safety-net setting. J Health Commun 2016;21(1):85-91 [FREE Full text] [doi: 10.1080/10810730.2015.1033118] [Medline: 26332306]
45. Haun JN, Lind JD, Shimada SL, Martin TL, Gosline RM, Antinori N, et al. Evaluating user experiences of the secure messaging tool on the Veterans Affairs' patient portal system. J Med Internet Res 2014;16(3):e75 [FREE Full text] [doi: 10.2196/jmir.2976] [Medline: 24610454]
46. Shimada SL, Petrakis BA, Rothendler JA, Zirkle M, Zhao S, Feng H, et al. An analysis of patient-provider secure messaging at two Veterans Health Administration medical centers: message content and resolution through secure messaging. J Am Med Inform Assoc 2017 Sep 01;24(5):942-949. [doi: 10.1093/jamia/ocx021] [Medline: 28371896]

47. US Department of Veterans Affairs. 2016. Connected Health Media Resources URL: https://www.va.gov/health/connectedhealth/index.asp [accessed 2017-08-09] [WebCite Cache ID 6savA3gSG]
48. Klein S. Commonwealth Fund. 2011. The Veterans Health Administration: Implementing Patient-Centered Medical Homes in the nation's largest integrated delivery system URL: http://www.commonwealthfund.org/Publications/Case-Studies/2011/Sep/VA-Medical-Homes.aspx [accessed 2017-08-03] [WebCite Cache ID 6sRhn34fw]
49. Rosland AM, Nelson K, Sun H, Dolan ED, Maynard C, Bryson C, et al. The patient-centered medical home in the Veterans Health Administration. Am J Manag Care 2013 Jul 01;19(7):e263-e272 [FREE Full text] [Medline: 23919446]
50. Werner RM, Canamucio A, Shea JA, True G. The medical home transformation in the Veterans Health Administration: an evaluation of early changes in primary care delivery. Health Serv Res 2014 Aug;49(4):1329-1347 [FREE Full text] [doi: 10.1111/1475-6773.12155] [Medline: 24476097]
51. Yano EM, Bair MJ, Carrasquillo O, Krein SL, Rubenstein LV. Patient Aligned Care Teams (PACT): VA's journey to implement patient-centered medical homes. J Gen Intern Med 2014 Jul;29 Suppl 2:S547-S549 [FREE Full text] [doi: 10.1007/s11606-014-2835-8] [Medline: 24715407]
52. Fix GM, Asch SM, Saifu HN, Fletcher MD, Gifford AL, Bokhour BG. Delivering PACT-principled care: are specialty care patients being left behind? J Gen Intern Med 2014 Jul;29 Suppl 2:S695-S702 [FREE Full text] [doi: 10.1007/s11606-013-2677-9] [Medline: 24715390]
53. Robert Graham Center. Citizens Health Initiative. 2007. The patient centered medical home: history, seven core features, evidence and transformational change URL: https://www.citizenshealthinitiative.org/sites/citizenshealthinitiative.org/files/media/Medical_Home/PCMHEvidenceTransformationalChange.pdf [accessed 2018-08-03] [WebCite Cache ID 6sRidnnsS]
54. US Department of Veterans Affairs. 2016. Access to Care - PACT URL: https://www.patientcare.va.gov/primarycare/pact/Access.asp [accessed 2018-08-09] [WebCite Cache ID 6sajz4tux]
55. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960 Apr 01;20(1):37-46. [doi: 10.1177/001316446002000104]
56. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
57. US Department of Veterans Affairs. 2017. Rural Veterans - Office of Rural Health URL: https://www.ruralhealth.va.gov/aboutus/ruralvets.asp [accessed 2018-02-09] [WebCite Cache ID 6x5e4PFad]
58. Bodenheimer T. Coordinating care--a perilous journey through the health care system. N Engl J Med 2008 Mar 6;358(10):1064-1071. [doi: 10.1056/NEJMhpr0706165] [Medline: 18322289]
59. Roter D. The enduring and evolving nature of the patient-physician relationship. Patient Educ Couns 2000 Jan;39(1):5-15. [Medline: 11013543]
60. Institute of Medicine (US) Committee on Quality of Health Care in America. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, D.C: National Academies Press (US); 2001.
61. Ngwenyama OK, Lee AS. Communication richness in electronic mail: critical social theory and the contextuality of meaning. MIS Q 1997 Jun;21(2):145-167. [doi: 10.2307/249417]
62. Caughlin JP, Bute JJ, Donovan-Kicken E, Kosenko KA, Ramey ME, Brashers DE. Do message features influence reactions to HIV disclosures? a multiple-goals perspective. Health Commun 2009 Apr;24(3):270-283. [doi: 10.1080/10410230902806070] [Medline: 19415559]
63. Kane B, Sands DZ. Guidelines for the clinical use of electronic mail with patients. The AMIA Internet Working Group, Task Force on Guidelines for the Use of Clinic-Patient Electronic Mail. J Am Med Inform Assoc 1998;5(1):104-111 [FREE Full text] [Medline: 9452989]
64. Robertson J. 2004. Guidelines for physician-patient electronic communications URL: http://www.ama-assn.org/ama/pub/category/2386.html [accessed 2006-03-03] [WebCite Cache ID 1141422610201782]

Abbreviations

EHR: electronic health record
OPCC&CT: Office of Patient-Centered Care and Cultural Transformation
PACT: Patient Aligned Care Teams
PHR: personal health record
VA: Veterans Health Administration

Edited by G Eysenbach; submitted 23.08.17; peer-reviewed by J Whealin, S Woods, F North; comments to author 14.09.17; revised version received 09.02.18; accepted 11.02.18; published 08.03.18.

Please cite as:
Hogan TP, Luger TM, Volkman JE, Rocheleau M, Mueller N, Barker AM, Nazi KM, Houston TK, Bokhour BG
Patient Centeredness in Electronic Communication: Evaluation of Patient-to-Health Care Team Secure Messaging
J Med Internet Res 2018;20(3):e82
URL: http://www.jmir.org/2018/3/e82/
doi:10.2196/jmir.8801
PMID:29519774

Original Paper

Digital Transformation and Disruption of the Health Care Sector: Internet-Based Observational Study

Maximilian Herrmann[1,2*], MSc; Philip Boehme[1,2*], Dr rer medic; Thomas Mondritzki[1,2], MSc; Jan P Ehlers[1], MA, DVM, PhD; Stylianos Kavadias[3], PhD; Hubert Truebel[2,4], MD, PhD

[1]Didactics and Educational Research in Health Science, Faculty of Health, Witten/Herdecke University, Witten, Germany

[2]Cardiovascular Research, Bayer Aktiengesellschaft, Wuppertal, Germany

[3]Judge Business School, University of Cambridge, Cambridge, United Kingdom

[4]Faculty of Health, Witten/Herdecke University, Witten, Germany

[*]these authors contributed equally

Corresponding Author:
Hubert Truebel, MD, PhD
Cardiovascular Research
Bayer Aktiengesellschaft
Apratherweg 18
Wuppertal,
Germany
Phone: 49 202360
Fax: 49 202364115
Email: hubert.truebel@bayer.com

Abstract

Background: Digital innovation, introduced across many industries, is a strong force of transformation. Some industries have seen faster transformation, whereas the health care sector only recently came into focus. A context where digital corporations move into health care, payers strive to keep rising costs at bay, and longer-living patients desire continuously improved quality of care points to a digital and value-based transformation with drastic implications for the health care sector.

Objective: We tried to operationalize the discussion within the health care sector around digital and disruptive innovation to identify what type of technological enablers, business models, and value networks seem to be emerging from different groups of innovators with respect to their digital transformational efforts.

Methods: From the Forbes 2000 and CBinsights databases, we identified 100 leading technology, life science, and start-up companies active in the health care sector. Further analysis identified projects from these companies within a digital context that were subsequently evaluated using the following criteria: delivery of patient value, presence of a comprehensive and distinctive underlying business model, solutions provided, and customer needs addressed.

Results: Our methodological approach recorded more than 400 projects and collaborations. We identified patterns that show established corporations rely more on incremental innovation that supports their current business models, while start-ups engage their flexibility to explore new market segments with notable transformations of established business models. Thereby, start-ups offer higher promises of disruptive innovation. Additionally, start-ups offer more diversified value propositions addressing broader areas of the health care sector.

Conclusions: Digital transformation is an opportunity to accelerate health care performance by lowering cost and improving quality of care. At an economic scale, business models can be strengthened and disruptive innovation models enabled. Corporations should look for collaborations with start-up companies to keep investment costs at bay and off the balance sheet. At the same time, the regulatory knowledge of established corporations might help start-ups to kick off digital disruption in the health care sector.

(J Med Internet Res 2018;20(3):e104) doi:10.2196/jmir.9498

KEYWORDS

digital transformation; health care sector; health care reform; incremental innovation; disruptive innovation; organizational innovation; entrepreneurship; efficiency; models; organizational; diffusion of innovation; delivery of health care

Introduction

Digital transformation and disruptive innovation describe the comprehensive reorientation of an industry including its business models due to the coming of age of digital technologies: the digitization of products, services, and processes [1-4]. It is expected that digital transformation of the health care sector will be as disruptive as that seen already in other industries [1-3,5]. Despite new technologies being constantly introduced, this change has yet to materialize [6-9].

According to Christensen [10,11], disruptive innovation requires 3 elements: (1) a technological enabler that simplifies previously complicated tasks, (2) a business model innovation that profitably delivers these simplified tasks in an affordable and convenient way, and (3) a new value network that reinforces a stakeholder position in this ecosystem. Given these conditions, it becomes intuitive that often enough disruptors come from outside an industry (with a total rethinking of the current business practices). They encroach the existing market dominance of established players from the bottom up (ie, from segments and products or services that can be viewed as lower margin and perhaps less valuable for the incumbent corporations in the industry) [9]. The counterpart to disruptive innovation is incremental innovation, the improvement or enhancement of product features and services that already exist in a market [6,12,13].

Health care systems face major challenges with rising costs, increasing demand for provision of care in aging societies, and outcome problems [14,15]. It has been shown in the United States that despite the availabilty of high-tech medicine, the average standard of care remains low compared to its cost [14], and this phenomenon can also be seen on a global scale [9,14,16].

Recent examples show that digital technology can mitigate or even eliminate these challenges, thus improving health care delivery [17-20]. Despite all the hype of "digital," why is the digital transformation of the health care sector still to be seen? One hurdle could be the heavily regulated nature of the sector. On one hand, regulations ensure that products reach the market with adequate safety, quality, and efficacy; on the other, regulating a complex industry could cause an innovation straightjacket because it is hard to predict the feasibility of innovative approaches well in advance [9,14,21,22]. For many patients, for whom health care remains expensive and at times inaccessible, the digital transformation offers the promise of better and cheaper care [6,10].

This study aims to provide an up-to-date comprehensive analysis of the transformational forces within the health care sector by looking at different stakeholders (life sciences, technology, and start-up companies). We evaluate their strategies on digital offerings and identify those that are disruptive or more incremental. We also point toward strategies that could enable digital disruption within the health care system.

Methods

Data

A systematic analysis was performed to screen for different technology and life science corporations regarding their digital transformation activities in the health care sector using 2017 Forbes 2000 data [23] from an annual ranking of the top 2000 companies in the world. The search terms "digital health," "digital medicine," "eHealth," "health care," "mHealth," "outcomes-based reimbursement," and "value-based care" were used to identify the 100 leading corporations. In addition, the 100 most successful start-up ventures active in the health care sector were identified based on the amount of funding they received as recorded from 2017 data by CBinsights [24]. We defined these efforts as "projects".

Evaluation of Identified Projects: Business Models, Solutions Provided, and Customer Needs Addressed

An expert panel consisting of 10 members with multiprofessional backgrounds in medicine, pharma, and economics rated these projects according to the following criteria [25,26]:

- Customer value proposition can be identified.
- Key resources can be identified.
- Key processes can be identified.
- Profit formula can be identified.

Each criterion was ranked from 0 (customer value proposition not given or not clear from the available sources) to 4 (customer value proposition can be readily identified). The last 3 criteria specifically allowed an assessment of the underlying business model [25,26]. The sums of scores from the 4 criteria were used to rank the projects and further look into the 20 highest ranked projects per group of companies (60 projects in total) in greater detail with regard to their regional location, customer value proposition, and solutions provided in connection with service, software, hardware, or platform to define different categories. These 60 projects were then evaluated according to 6 customer needs: adherence, diagnostic, lifestyle, patient engagement, prevention, or treatment. These categories were created by identifying similarities between the different projects and grouping them by which customer needs they addressed. The groups were named accordingly. A chi-square test was performed for the different companies to verify whether their provided solutions or their addressed customer needs come from the same distribution or are significantly different. All tests were performed with statistical significance of $P<.05$.

Results

Regional Distribution

More than 400 projects (Multimedia Appendix 1) were identified from the 100 leading start-up, life science, and technology companies. In our analysis of the 60 highest rated projects (Multimedia Appendix 2) identified by our expert panel, a high regional concentration with 40 out of 60 projects (66%) located in the United States was found. There was an aggregation of

projects on the West Coast of the United States (Multimedia Appendix 3).

Business Models, Solutions Provided, and Customer Needs Addressed

Multimedia Appendix 4 shows the results of our in-depth analysis of the customer value proposition and the underlying business model. Within the group of highest ranked projects, device-developing start-up projects assume a primary position. The second grouping represents different start-up projects that are less well-defined with respect to their customer value proposition and business model. In the third grouping, one can see that corporations from the technology and life science sectors appear to be more active in collaborations and in efforts to engage with topical experts in their targeted digital innovations.

Provided Solutions

The analysis led to 4 distinct types of solutions that could be identified, as shown in Figure 1. All 3 industry players are engaging in projects that are represented across the different types of solutions. Specifically, while hardware solutions appear to be evenly distributed, projects on new services seem to be undertaken only by start-ups and life sciences companies. Interestingly we found a strong engagement of technology corporations in the platform field. Chi-square tests indicated there is no statistical evidence that start-up and technology solutions (χ^2_3=21.2, P<.001,), start-up and life science solutions (χ^2_3=26.9, P<.001), or technology and life science solutions (χ^2_3=30.2, P<.001) come from the same distribution. This indicates with respect to the structure of the solutions that the 3 groups have been pursuing distinctively different strategies.

Customer Needs Addressed

To analyze whether these projects were addressing similar or different customer needs, we focused on 6 customer needs that were further investigated: adherence, diagnostic, lifestyle, patient engagement, prevention, and treatment. As shown in Figure 2, distributions among the different companies' foci were found. It could be shown that the start-up projects represent all patient needs consistently. Efforts from the life science sector were focused primarily on adherence- and treatment-related projects, while no major actions appeared for prevention, diagnostic, and lifestyle. Technology corporations were similar with no projects in the lifestyle field. Chi-square tests indicated there is no statistical evidence that start-up– and technology-addressed needs (χ^2_5=60.5, P<.001) or start-up– and life science–addressed needs (χ^2_5=85.3, P<.001) come from the same distribution. However chi-square tests indicated there is statistical evidence that technology- and life science–addressed needs come from the same distribution (χ^2_4=3.8, P=.435). That means that start-up companies have a significantly different focus than technology and life sciences corporations. However, the technology and life sciences corporations do not have statistically significant foci from each other.

Intercategorial View

Combining the information from the previous analyses shows that adherence and treatment projects are preferred by corporations in the life science and technology sectors (Figure 3). Both seem to neglect lifestyle-focused projects. More importantly, start-up companies show a much broader focus in their efforts to address patient needs.

Figure 1. Distribution of identified types of solutions among the projects of the 3 industry players.

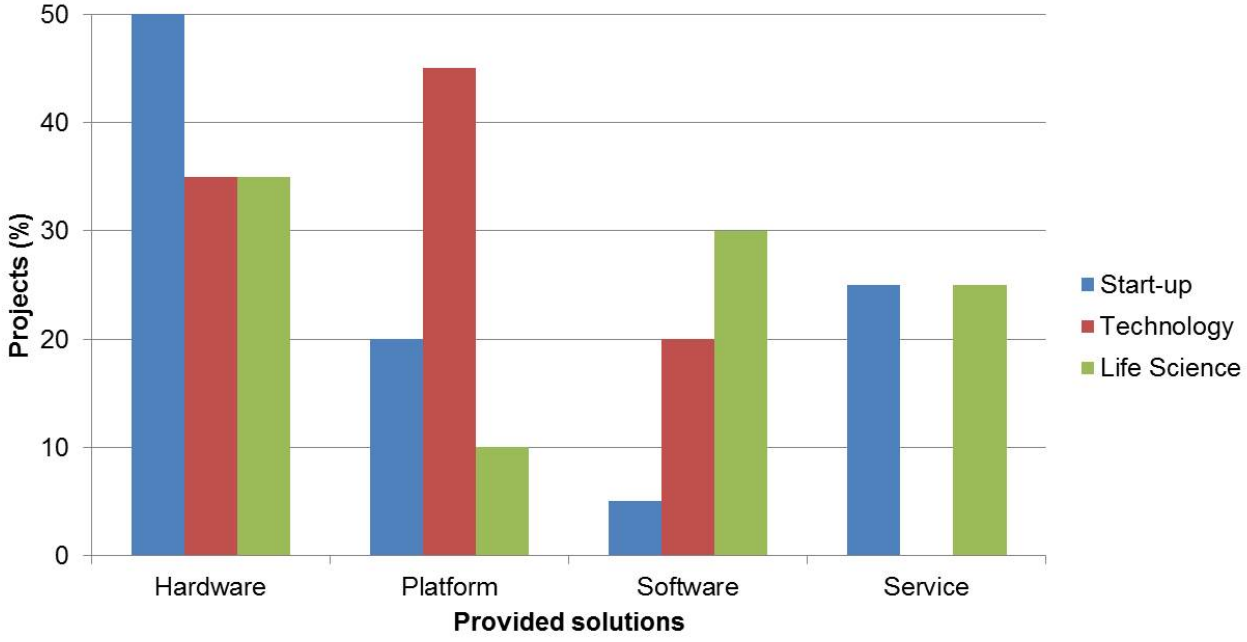


Figure 2. Distribution of customer needs addressed among the projects of the 3 industry players.

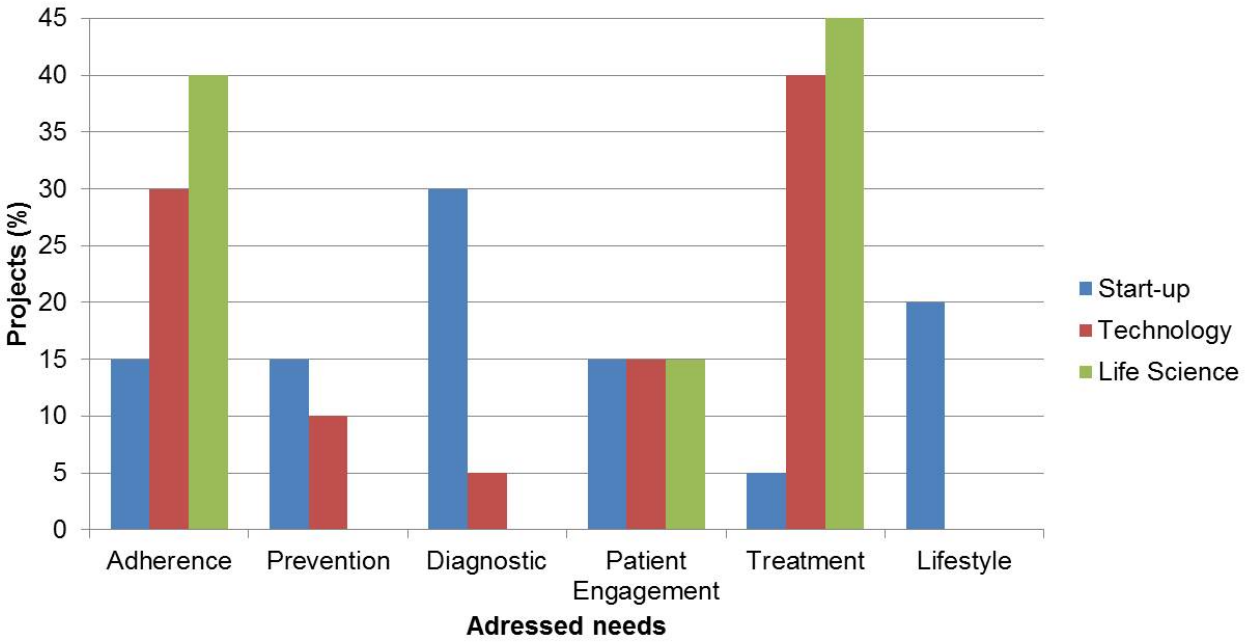
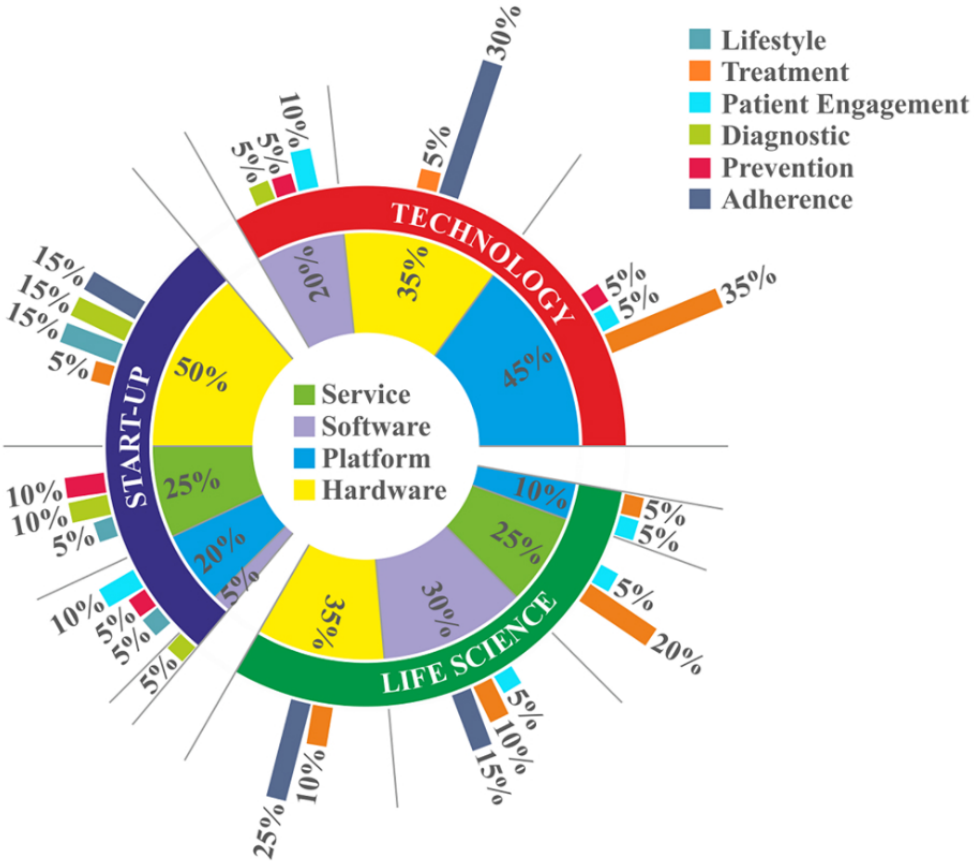


Figure 3. Solutions provided and needs addressed by category.



Discussion

Summary

The goal of the study was to provide an up-to-date comprehensive analysis of the transformational forces within the health care sector by looking at different stakeholders. Our results identified patterns showing that established corporations rely more on incremental innovation that supports their current business model, while start-up companies engage their flexibility to explore new market segments with notable transformations of established business models.

Textbox 1. Example of the focus of a life science company.

Sanofi is a French-based life science company that has signed a value-based pricing contract with the US health insurer Cigna on a new cholesterol-lowering drug. The price is linked to patient cholesterol levels. If the drug fails to decrease the level of cholesterol as seen in clinical trials, Sanofi must further discount the drug.

Textbox 2. Example of a not fully developed value proposition and business model.

xbird is a German-based start-up company that extracts millions of data points from smartphone sensors, wearables, and medical devices, combining environmental and digital biomarkers. Data scientists and medical experts combine these data and identify patterns leading to critical health events. This technology enables connection of adverse health events with behavioral causes and creates actionable insights for both doctors and patients. The company implements technology, new business models, and value propositions to use the collected data to avoid critical health events before occurrence.

Regional Diversity

The data show a dominant positioning of projects in the United States. One reason could be that the United States shows the highest digital innovation potential through the provision of an extraordinary environment for disruptive innovation. Here our analysis indicates that 17 projects out of the 40 (43%) were located in California, and 10 out of the 17 projects (59%) are based in San Francisco and Mountain View, the heart of Silicon Valley. There might also be an increased interest for disruptive innovation in the health care sector in the United States due to its low efficiency [10].

Principal Findings

The distribution of business models provided by the different projects indicates 3 main areas that distinguish large, established corporations (life science and technology) from start-ups. The younger start-up competitors appear to pursue solo efforts instead of collaborative endeavors and undertake efforts directed in spaces that are not pursued by others. When exploring further into the detailed solutions (Figure 1) and addressed needs (Figure 2), we find that start-ups display a wider approach toward the digital health care sector. This is contrasted by the more established life science and technology corporations which focus on adherence and treatment projects (Textbox 1). These supplement their existing market offerings, and therefore could be viewed as initial departures from the existing business models. However, statistical results indicate that not a single cohort of companies has figured out what the right digital approach is and this reflects well in the traditional fermentation/converging period that many industries exhibit during large shifts in how business takes place.

With respect to the customer needs addressed, we notice an interesting effect. Technology and life science corporations seem to address similar customer needs. Either the technology companies have not been really creative in addressing the digital challenges or the technology companies were creative in the beginning but the life science sector has caught up pretty fast. The start-up endeavors are significantly different than all others, which indicates that disruption would come from the start-ups because they are playing in different domains than established corporations. In that regard, they exhibit strong differentiation in both the supply and the demand side from a digital innovation standpoint. These observations can echo the underlying structural elements in the disruptive innovation theory of Christensen [10,11]. In that light, incumbent corporations within the life science sector (eg, pharmaceutical corporations) tend to work on more effective drugs, but they lack the capability to directly interact with the patient and therefore transform their competitive position by additional (recently termed “beyond the pill”) offerings. Aligned with this expectation, we identified that their projects show a clear tendency to offer digitally enhanced outputs but such efforts tend to be incremental innovations that stick to traditional market strategies. Interestingly, providers of consumer care products like Fitbit and Jawbone are further penetrating the health care sector, moving beyond lifestyle products for customers interested in health self-monitoring and toward offerings that compete directly with more established health care corporations like Medtronic.

Start-up offerings of products and services appear in some cases not fully developed with respect to the exact value proposition of the offering (Multimedia Appendix 4) and business model (Textbox 2).

In general, start-ups use their flexible structures to pursue radically new avenues with the help of novel technologies, business models, and value networks that provide disruptive solutions to a wide variation of customer and patient challenges. Within our sample, established technology and life science corporations aim through their projects to address challenges that relate, to a significant extent, to the adherence and treatment dimensions of the customer value. Thus, they seem to be underrepresented in the remaining types of customer value. This offers evidence that established corporations focus on digital improvements of their existing business offerings and value proposition, which in turn signifies lesser interventions to their current business models. Their focus stands in sharp contrast to the diversity of start-ups, which seem to address diversified customer needs.

Limitations

Our analysis, as the first capturing the phenomenon of disruptive innovation within the health care sector, has a few limitations. It is based on information available in the public domain, which might not allow for a comprehensive picture since some start-ups might overreport to attract funding and other start-ups might not have yet made a public splash, as they are rather early in their development process. At the same time, established large corporations for privacy reasons might underreport on their digital initiatives. The ratings by 10 evaluators might also blur stronger differences given their diverse educational backgrounds. Public domain data might lack the depth of information needed

to allow for precise rating of the different characteristics assessed.

Conclusion

Notwithstanding the limitations of our analysis, the emerging pattern allows us to differentiate innovating corporations within the health care sector with respect to their strategies in the context of the digital transformation in health care. Established corporations show strength in improving the business model dimensions they have been pursuing for a long time. Start-up companies appear more agile and able to make better use of radical new technologies and different business models moving toward new forms of disruptive innovations. Since the health care sector is tightly regulated, established players with an in-depth understanding of its regulatory mechanism might have clear advantages here, but start-ups are tackling this specific challenge well.

Start-ups with their agile culture and established technology or life science corporations with their regulatory knowledge might join strategic forces to drive the digital transformation of the health care sector. By engaging in collaborative efforts, corporations can keep costs at bay, while addressing all patient needs and claiming the investments off their balance sheet. Being in position to quickly adapt when a disruptive business model emerges will be the key for future revenues. A disruptive threat for both life science and start-up is the strong focus of technology corporations to establish platform business models and assume the necessary bargaining power to appropriate the value created. It remains to be seen whether the future market leaders of a transformed health care sector will be the existing corporations and current market leaders or new players who are going to emerge from the ranks of today's start-ups.

Acknowledgments

The authors would like to thank the expert panel that helped rank the projects.

Conflicts of Interest

MH, PB, TM, and HT were full-time employees of Bayer Aktiengesellschaft at the time of preparation of the manuscript.

Multimedia Appendix 1

List of identified projects.

[PDF File (Adobe PDF File), 241KB - jmir_v20i3e104_app1.pdf]

Multimedia Appendix 2

The 60 highest ranked projects by expert evaluation.

[PDF File (Adobe PDF File), 264KB - jmir_v20i3e104_app2.pdf]

Multimedia Appendix 3

Regional diversity of the 60 highest ranked projects.

[JPG File, 77KB - jmir_v20i3e104_app3.jpg]

Multimedia Appendix 4

Distribution of projects.

[JPG File, 107KB - jmir_v20i3e104_app4.jpg]

References

1. Harvard Business Review. 2014 Oct 01. Digital transformation in the high-tech industry: briefing paper sponsored by SAP URL: https://hbr.org/resources/pdfs/comm/sap/18764_HBR_SAP_High_Tech_Aug_14.pdf [accessed 2017-11-08] [WebCite Cache ID 6uoiUFoim]
2. Kreutzer R. Treiber und Hintergrunde der digitalen Transformation. In: Schallmo D, Rusnjak A, Anzengruber J, Werani T, Junger M, editors. Digitale Transformation von Geschäftsmodellen: Grundlagen, Instrumente und Best Practices. Wiesbaden: Springer Fachmedien; 2017:33-58.
3. Valenduc G, Vendramin P. Digitalisation, between disruption and evolution. Transfer Eur Rev Labour Res 2017 Apr 13;23(2):121-134. [doi: 10.1177/1024258917701379]
4. Loucks J, Macaulay J, Norohna A, Wade M. Digital Vortex: How Today's Market Leaders Can Beat Disruptive Competitors at Their Own Game. Plano: DBT Center Press; 2016.

5. Klewes J, Popp D, Rost-Hein M. Out-thinking Organizational Communications: The Impact of Digital Transformation. Basel: Springer International Publishing; 2016.
6. Hwang J, Christensen CM. Disruptive innovation in health care delivery: a framework for business-model innovation. Health Aff (Millwood) 2008;27(5):1329-1335 [FREE Full text] [doi: 10.1377/hlthaff.27.5.1329] [Medline: 18780919]
7. Perakslis ED. Strategies for delivering value from digital technology transformation. Nat Rev Drug Discov 2017 Feb;16(2):71-72. [doi: 10.1038/nrd.2016.265] [Medline: 28082744]
8. Westerman G, Tannou M, Bonnet D, Ferraris P, McAfee A. The digital advantage: how digital leaders outperform their peers in every industry. URL: http://digitalcommunity.mit.edu/servlet/JiveServlet/downloadBody/1105-102-2-1185/The_Digital_Advantage__How_Digital_Leaders_Outperform_their_Peers_in_Every_Industry.pdf [accessed 2017-11-08] [WebCite Cache ID 6uoktbLY4]
9. Porter ME, Teisberg EO. Redefining competition in health care. Harv Bus Rev 2004 Jun;82(6):64-76. [Medline: 15202288]
10. Christensen C, Grossman J, Hwang M. The Innovator's Prescription: A Disruptive Solution for Health Care. New York: McGraw-Hill; 2008.
11. Christensen C. The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail. Watertown: Harvard Business Review Press; 2013.
12. Hacklin F, Raurich V, Marxt C. How incremental innovation becomes disruptive: the case of technology convergence. 2004 Presented at: IEEE International Engineering Management Conference; 2004; Singapore p. 18-21. [doi: 10.1109/IEMC.2004.1407070]
13. Assink M. Inhibitors of disruptive innovation capability: a conceptual model. Euro J Innov Managem 2006 Apr;9(2):215-233. [doi: 10.1108/14601060610663587]
14. Porter M, Teisberg E. Redefining Health Care: Creating Value-Based Competition on Results. Watertown: Harvard Business School Press; 2006.
15. Christensen CM, Bohmer R, Kenagy J. Will disruptive innovations cure health care? Harv Bus Rev 2000;78(5):102-112. [Medline: 11143147]
16. Kohn LT, Corrigan JM, Donaldson MS, editors. To Err is Human: Building a Safer Health System. Washington: National Academies Press; 2000.
17. Labovitz DL, Shafner L, Reyes GM, Virmani D, Hanina A. Using artificial intelligence to reduce the risk of nonadherence in patients on anticoagulation therapy. Stroke 2017 May;48(5):1416-1419. [doi: 10.1161/STROKEAHA.116.016281] [Medline: 28386037]
18. 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 May 26;19(5):e177 [FREE Full text] [doi: 10.2196/jmir.7374] [Medline: 28550006]
19. Steinhubl SR, Muse ED, Topol EJ. Can mobile health technologies transform health care? JAMA 2013 Dec 11;310(22):2395-2396. [doi: 10.1001/jama.2013.281078] [Medline: 24158428]
20. Kaufman N, Salahi A. Using digital health technology to prevent and treat diabetes. Diabetes Technol Ther 2017 Feb;19(S1):S59-S73. [doi: 10.1089/dia.2017.2506] [Medline: 28192015]
21. Goddard M. Competition in healthcare: good, bad or ugly? Int J Health Policy Manag 2015 Aug 01;4(9):567-569 [FREE Full text] [doi: 10.15171/ijhpm.2015.144] [Medline: 26340484]
22. Mwachofi A, Al-Assaf AF. Health care market deviations from the ideal market. Sultan Qaboos Univ Med J 2011 Aug;11(3):328-337 [FREE Full text] [Medline: 22087373]
23. Forbes 2000 list. New York City: Forbes URL: https://www.forbes.com/global2000/list/ [accessed 2018-03-03] [WebCite Cache ID 6xcid8HC8]
24. CBinsights. New York City: CBinsights URL: https://www.cbinsights.com/ [accessed 2018-03-03] [WebCite Cache ID 6xcjQLQvc]
25. Johnson M, Lafley A. Seizing the White Space: Business Model Innovation for Growth and Renewal. Watertown: Harvard Business Review Press; 2010.
26. Johnson MW, Christensen CM, Kagermann H. Reinventing your business model. Harvard Bus Rev 2008;86(12):57-68.

Edited by G Eysenbach; submitted 26.11.17; peer-reviewed by K Agarwal, R Pankomera; comments to author 17.01.18; revised version received 26.01.18; accepted 26.01.18; published 27.03.18.

Please cite as:
Herrmann M, Boehme P, Mondritzki T, Ehlers JP, Kavadias S, Truebel H
Digital Transformation and Disruption of the Health Care Sector: Internet-Based Observational Study
J Med Internet Res 2018;20(3):e104
URL: http://www.jmir.org/2018/3/e104/
doi:10.2196/jmir.9498
PMID:29588274

Letter to the Editor

Comment on: Clinical Validity, Understandability, and Actionability of Online Cardiovascular Disease Risk Calculators: Systematic Review

Ivan Sisa[1], MD, MPH, MS

School of Medicine, College of Health Sciences, Universidad San Francisco de Quito, Quito, Ecuador

Corresponding Author:
Ivan Sisa, MD, MPH, MS
School of Medicine
College of Health Sciences
Universidad San Francisco de Quito
Diego de Robles y Via Interoceanica
Quito,
Ecuador
Phone: 593 2 297 1700 ext 4017
Email: isisa@usfq.edu.ec

Related Article:

Comment on: http://www.jmir.org/2018/2/e29/

(J Med Internet Res 2018;20(3):e10093) doi:10.2196/10093

KEYWORDS

cardiovascular disease; risk assessment; risk model

I am writing regarding the systematic review about clinical validity, understandability, and actionability of online cardiovascular disease (CVD) risk calculators recently published by Dr Bonner and colleagues [1].

Although Dr Bonner and colleagues used a comprehensive two-step research strategy to identify Web addresses that contained a CVD risk calculator, which led to the identification of 67 Web pages, a very important CVD risk model, the Systematic COronary Risk Evaluation (SCORE) risk assessment model [2], was ignored. Developed by the European Society of Cardiology, this model was derived from 12 European cohort studies (250,000 patients data collected and 3 million person-years of observation) and is based on classical risk factors such as gender, age, total cholesterol, systolic blood pressure, and smoking status. The SCORE risk assessment model should have been included because it satisfies Dr Bonner's inclusion criteria as it predicts the risk of developing a CVD event and an electronic interactive version of this model is freely available on the European Society of Cardiology's Web page [3].

Furthermore, there are other risk assessment models locally developed in countries such as China, India, and Korea that are not taken into account in this study. Thus, it would have been useful if the authors had added to their research strategy a literature search of review documents focusing on cardiovascular risk assessment as was carried out by Zhao and colleagues [4].

Editorial note: Authors were invited to respond but declined. They agree that additional calculators exist that could have been included if a different method was used.

References

1. Bonner C, Fajardo MA, Hui S, Stubbs R, Trevena L. Clinical Validity, Understandability, and Actionability of Online Cardiovascular Disease Risk Calculators: Systematic Review. J Med Internet Res 2018 Feb 01;20(2):e29 [FREE Full text] [doi: 10.2196/jmir.8538] [Medline: 29391344]
2. Conroy RM, Pyörälä K, Fitzgerald AP, Sans S, Menotti A, De BG, SCORE project group. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. Eur Heart J 2003 Jun;24(11):987-1003. [Medline: 12788299]

3. European Society of Cardiology. SCORE risk charts: the European cardiovascular disease risk assessment model URL: https://www.escardio.org/Education/Practice-Tools/CVD-prevention-toolbox/SCORE-Risk-Charts [accessed 2018-02-10]
4. Zhao D, Liu J, Xie W, Qi Y. Cardiovascular risk assessment: A global perspective. Nat Rev Cardiol 2015;12:301-311. [doi: 10.1038/nrcardio.2015.28]

Abbreviations

CVD: cardiovascular disease
SCORE: Systematic COronary Risk Evaluation

Edited by G Eysenbach; submitted 11.02.18; this is a non–peer-reviewed article;accepted 14.02.18; published 05.03.18.

Please cite as:
Sisa I
Comment on: Clinical Validity, Understandability, and Actionability of Online Cardiovascular Disease Risk Calculators: Systematic Review
J Med Internet Res 2018;20(3):e10093
URL: http://www.jmir.org/2018/3/e10093/
doi:10.2196/10093
PMID:29506971

Corrigenda and Addenda

Figure Caption Correction: Characteristics of Articles About Human Papillomavirus Vaccination in Japanese Newspapers: Time-Series Analysis Study

Nao Ueda[1], BA; Ryoki Yokouchi[2*], BA; Taro Onoda[3*], BS (Med); Atsushi Ogihara[4*], PhD

[1]Graduate School of Human Sciences, Waseda University, Tokorozawa, Japan

[2]Department of Health Sciences and Social Welfare, School of Human Sciences, Waseda University, Tokorozawa, Japan

[3]Department of Medicine, School of Medicine, Tokai University, Isehara, Japan

[4]Faculty of Human Sciences, Waseda University, Tokorozawa, Japan

[*]these authors contributed equally

Corresponding Author:
Nao Ueda, BA
Graduate School of Human Sciences
Waseda University
2-579-15
Mikajima
Tokorozawa, 359-1192
Japan
Phone: 81 80 5082 3970
Email: naaaaa0428@gmail.com

Related Article:

Do not cite this article - this article has been retracted. Retraction statement: http://publichealth.jmir.org/2018/2/e10878/

(J Med Internet Res 2018;20(3):e27) doi:10.2196/jmir.9728

This correction notice was retracted on May 10, 2018, as it was mistakenly published in J Med Internet Res rather than JMIR Public Health Surveill. For details, please see the retraction statement included in the new correction notice (JMIR Public Health Surveill 2018;4(2):e10878).

Edited by G Eysenbach; submitted 24.12.17; this is a non–peer-reviewed article;accepted 10.01.18; published 16.03.18.

Please cite as:
Ueda N, Yokouchi R, Onoda T, Ogihara A
Figure Caption Correction: Characteristics of Articles About Human Papillomavirus Vaccination in Japanese Newspapers: Time-Series Analysis Study
J Med Internet Res 2018;20(3):e27
URL: http://www.jmir.org/2018/3/e27/
doi:10.2196/jmir.9728
PMID:29547387

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org