Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 20 (2018), Issue 10 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Original Papers

Reviews

Viewpoints

Tutorial

Letters to the Editor

Original Paper

How Users Experience and Use an eHealth Intervention Based on Self-Regulation: Mixed-Methods Study

Louise Poppe[1,2], MSc; Celien Van der Mispel[1,2], MSc; Geert Crombez[2], PhD; Ilse De Bourdeaudhuij[1], PhD; Helene Schroé[1,2], MSc; Maïté Verloigne[1], PhD

[1]Research Group Physical Activity and Health, Department of Movement and Sports Sciences, Ghent University, Ghent, Belgium
[2]Ghent Health Psychology Lab, Department of Experimental-Clinical and Health Psychology, Ghent University, Ghent, Belgium

Corresponding Author:
Louise Poppe, MSc
Ghent Health Psychology Lab
Department of Experimental-Clinical and Health Psychology
Ghent University
Henri Dunantlaan 2
Ghent, 9000
Belgium
Phone: 32 9 264 86 51
Email: louise.poppe@ugent.be

Abstract

Background: eHealth interventions show stronger effects when informed by solid behavioral change theories; for example, self-regulation models supporting people in translating vague intentions to specific actions have shown to be effective in altering health behaviors. Although these theories inform developers about which behavioral change techniques should be included, they provide limited information about how these techniques can be engagingly implemented in Web-based interventions. Considering the high levels of attrition in eHealth, investigating users' experience about the implementation of behavior change techniques might be a fruitful avenue.

Objective: The objective of our study was to investigate how users experience the implementation of self-regulation techniques in a Web-based intervention targeting physical activity and sedentary behavior in the general population.

Methods: In this study, 20 adults from the general population used the intervention for 5 weeks. Users' website data were explored, and semistructured interviews with each of the users were performed. A directed content analysis was performed using NVivo Software.

Results: The techniques "providing feedback on performance," "action planning," and "prompting review of behavioral goals" were appreciated by users. However, the implementation of "barrier identification/problem solving" appeared to frustrate users; this was also reflected by the users' website data—many coping plans were of poor quality. Most users were well aware of the benefits of adopting a more active way of living and stated not to have learned novel information. However, they appreciated the provided information because it reminded them about the importance of having an active lifestyle. Furthermore, prompting users to self-monitor their behavioral change was not sufficiently stimulating to make users actually monitor their behavior.

Conclusions: Iteratively involving potential end users offers guidance to optimally adapt the implementation of various behavior change techniques to the target population. We recommend creating short interventions with a straightforward layout that support users in creating and evaluating specific plans for action.

(J Med Internet Res 2018;20(10):e10412) doi:10.2196/10412

KEYWORDS

eHealth; self-regulation; behavioral change theory; interview; usage data

Introduction

eHealth, or "the use of technology to improve health care" [1] is effective in changing health behaviors, such as increasing physical activity, altering dietary habits, and smoking cessation [2-4]. Furthermore, eHealth programs have the potential to reach large populations in a cost-effective way [5-7]. They may also enable a personalized and interactive approach, for example, by computer tailoring, without the practical considerations of face-to-face contacts [7-9].

There are strong indications that eHealth interventions should be informed by sound theories. Research has shown that applying a theoretical basis to eHealth interventions increases their effectiveness [10,11]; for example, self-regulation models [12] have identified several techniques that may help users to engage in behavioral change. Self-regulation is the process of goal selection, pursuit, and maintenance [13]; it focuses not only on eliciting an intention to change behavior but also on bridging the gap between intention and behavior [13-15]. Using the self-regulation perspective, individuals may learn how to initiate change effectively and how to maintain health behavior over changing conditions. “Action planning,” for example, comprises the detailed planning of what a person will do, whereas “barrier identification/problem solving” helps individuals to identify and solve difficult situations for performing the health behavior [16]. Furthermore, research has shown that self-regulation strategies are, indeed, effective in changing behavior [17-21].

Although behavioral change theories inform us about which behavioral change techniques should be included, they provide limited information about how these techniques can be implemented in an engaging way [10]; this might explain why Web-based and mobile interventions often suffer from high attrition rates (60%-80%) [22-24]. The use of behavioral change theories may be necessary but not sufficient to guarantee efficacious interventions. Equally important is the involvement of potential users during various stages of the development process. Such an approach has been advocated by many and is known as cocreation [25], person-based approach [26], or user-centered development [27].

Involving the target population has given researchers insight into what motivates users to start and adhere to a Web-based intervention; for example, Bardus et al. found that the expectation of receiving reminders regarding physical activity was an important reason to start with a Web-based physical activity intervention [28]. Time efficiency, a clear navigation structure, and professional design of the eHealth intervention have been shown to be important factors to make users stay in the program [29,30]. Finally, providing users with a sense of control motivates them to complete the eHealth program [31]. These findings act as a guide to further fine-tune eHealth interventions to the target population [26].

This study aims to investigate how users experience self-regulation techniques implemented in an eHealth intervention. For this purpose, we used the eHealth intervention “MyPlan 2.0,” which supports users to be more physically active or less sedentary in a step-by-step manner. This intervention is informed using self-regulation theory and considers users as their own expert in the behavioral change process. Through a semistructured interview and an examination of users’ website data, information was obtained about the appreciation of the website and intervention in general and the experience of users with various self-regulation techniques (ie, goal setting, providing information, providing feedback on performance, action planning, barrier identification/problem solving, prompting self-monitoring, planning social support, and reviewing behavioral goals). The findings derived from this study might help other eHealth developers on how (not) to implement self-regulation techniques in Web-based interventions.

Methods

Participants

In this study, 20 adults from the general population volunteered to participate; this number was based on previous qualitative research about eHealth by Yardley et al. [32]. Participants were recruited via acquaintances of the researchers and a database of the research group. The database contained the names of persons who expressed interest in participating in studies of the Ghent Health Psychology Research Group. The exclusion criteria were as follows: not having internet access, aged <18 years, diagnosed with a chronic disease, and non-Dutch speaking. To maintain an equal distribution over age, gender, and educational level, we preselected participants based on these characteristics. The study was conducted between November 2016 and May 2017. As soon as a participant was enrolled in the study, he or she could start the intervention. The first participant started in November 2016, and the last participant started in April 2017. The study was approved by the Committee of Medical Ethics of the Ghent University Hospital (Belgian registration number: B670201629995), and all participants provided a written informed consent.

Intervention

“MyPlan 2.0” is a self-regulation-based intervention consisting of 5 weekly Web-based sessions. It aims to increase physical activity and decrease sedentary behavior in adults and is designed and created by our research group. “MyPlan 2.0” is based on a previous version named “MyPlan 1.0” [33], which was effective in changing health behaviors [21,33-35]. However, the quantitative research with “MyPlan 1.0” revealed high levels of nonusage attrition [36]. The qualitative research revealed that users felt frustrated about the length and complexity of the program [30]. Hence, the intervention was iteratively transformed according to this feedback. In particular, the intervention was shortened, the text was limited, information sheets were substituted by a quiz, and the layout was changed. Furthermore, rationales were provided for the implementation of different self-regulation techniques, specific instructions were given during action planning and barrier identification/problem solving, and general tips and tricks were provided. Moreover, success stories of other users were added.

In the first session, participants started by creating a profile and provided general information (eg, gender, age, and working status) to enable personalized messages during the intervention. In addition, they chose which behavior, physical activity or sedentary behavior, they wanted to change during the intervention (ie, “goal setting”). The website offers the option to take a quiz regarding the chosen health behavior (ie, “providing information on the consequences of the behavior”). Thereafter, participants completed a short questionnaire regarding the selected health behavior, that is, a shortened version of the International Physical Activity Questionnaire

(IPAQ) [37] or a last 7-days sedentary behavior questionnaire [38] and received tailored feedback, that is, "providing feedback on performance." After receiving feedback, participants were guided to the "action planning" technique. During this component, users specified their actions in terms of what, where, and how by answering open- and multiple-choice questions. Several tips were provided to make the action plan feasible (eg, "Choose for one goal instead of multiple goals, this increases the chance of goal attainment"). Next, "barrier identification/problem solving" was introduced by asking users which barriers they could perceive and which solutions were possible to overcome these barriers. In addition, examples of barriers and related solutions were provided, which could be selected by users. Next, "prompting self-monitoring of behavior" was introduced. Users chose from a list how they would monitor their behavior (eg, via their calendar, in a notebook, and so on). During the action planning, barrier identification/problem solving, and self-monitoring component, success stories from fictitious users were shown; these were incorporated to elicit motivation further and provide inspiration. At the end of the first session, "planning social support" was introduced; users read about how to elicit social support, how to talk about behavioral change to significant others, and how to find opportunities to engage in behavioral change together with other people. Figure 1 depicts the flowchart of the first session. Multimedia Appendix 1 shows the exact implementation of the techniques through screenshots.

After 1 week, users received an email to return to the eHealth program to revise their plan. According to the technique "Prompt review of behavioral goals," they were asked how well the behavioral change was going and whether they wanted to adapt or maintain their plan. If they wanted to adapt their plan, action planning was again completed. In all cases, users were prompted for barrier identification and problem solving. To motivate users to think about more personally relevant barriers and solutions, users now answered an open-ended question instead of selecting an option from a predefined list. A summary of their answers was shown in the action plan, and users were prompted to self-monitor their behavior. In addition, users could again read the information about social support and receive extra tips and tricks, and this illustrated the use of different self-regulation techniques, such as "prompting rewards," prompting focus on past success," "providing instructions," "teaching to use prompts/cues," and "prompting self-talk;" this cycle was the same for each of the 4 follow-up sessions. Figure 2 displays the flowchart of the follow-up sessions.

The effect of "MyPlan 2.0" will be tested by a randomized controlled trial. If the intervention is effective, it will be disseminated and implemented by the "Flemish Institute for Healthy Living," which is the Flemish center of expertise regarding health promotion and illness prevention.

Procedure

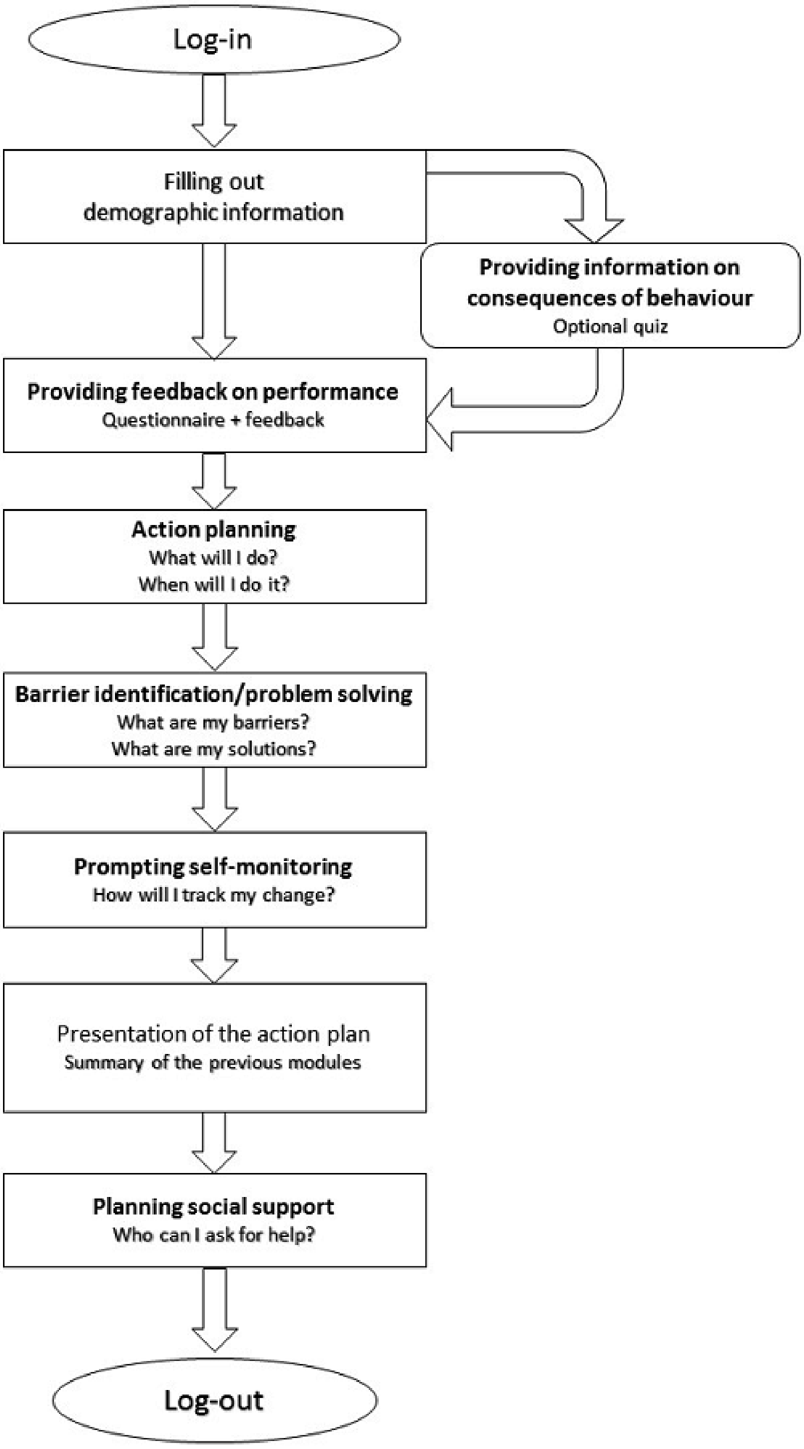
Participants were contacted by telephone and informed about the study. When participants decided to take part in the study, they received an email with a website link to the intervention and the documents to provide their informed consent. Participants were instructed to complete the intervention on their own. When researchers noted that participants forgot to log in at the scheduled time, they were reminded of doing so by a telephone call. After completing 5 intervention sessions, users' website data were downloaded, and a date to perform a semistructured interview was scheduled. Before the start of the interview, participants completed questions about demographic characteristics (ie, age, gender, educational level, height, and weight). The interviews took place at the research department or via a telephone call. The interviews were audiorecorded with permission of participants.

The questions and content of the semistructured interview were based on the results of the previous qualitative research with the intervention "MyPlan 1.0" [30]. The 3 main topics that were addressed during the interview were as follows: design of the intervention (ie, general appreciation, user-friendliness, time efficiency, and layout); usefulness of the website (ie, opinion about the motivational value of the website, opinion about the informative value, feelings of awareness elicited by the website, personal relevance, and recommendations); and views about the benefits of being more physically active or less sedentary. During the discussion of each topic, researchers explicitly focused on how users had experienced each of the self-regulation techniques implemented in "MyPlan 2.0" (eg, "How did you experience the component in which you were asked to formulate personal barriers and solutions?"). The interview guide can be found in Multimedia Appendix 2. In the Results section, we will focus on perceptions' regarding the website in general and the implementation of the behavioral change techniques. The average duration of an interview was 30 minutes, and participants received a reimbursement of €20.

Data Analysis

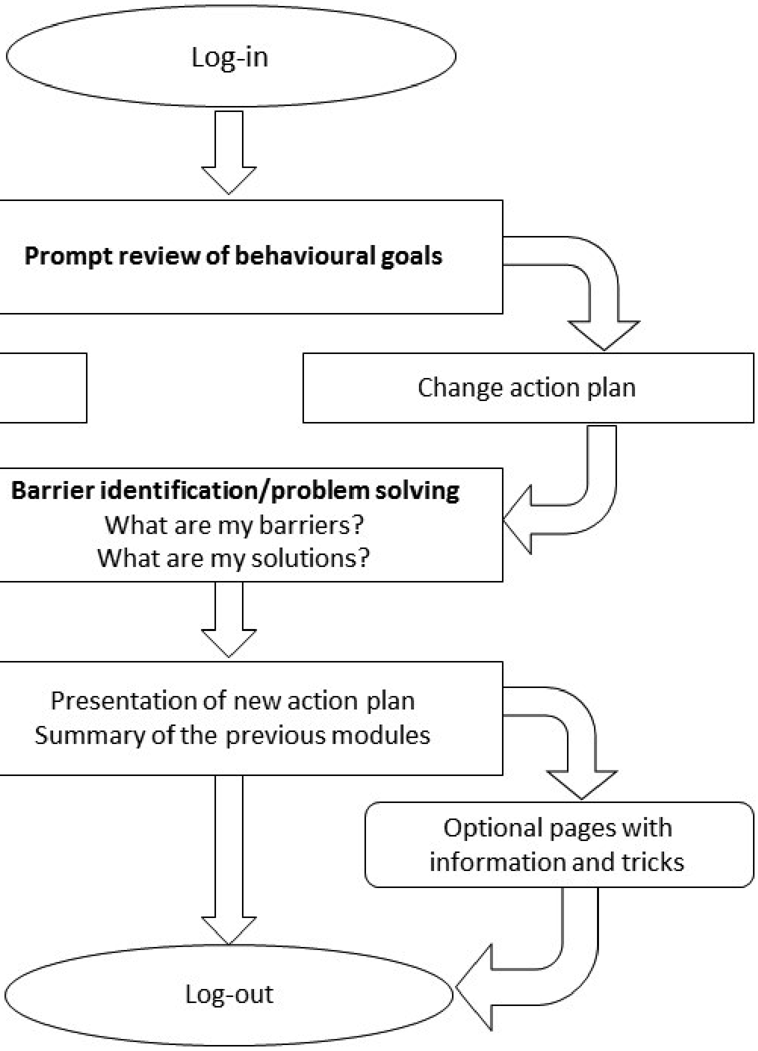
The following information was derived from the users' website data. First, we identified how many users selected sedentary behavior and physical activity as their target behavior and how many received the tailored feedback that they did not meet the respective health norm (ie, 30 minutes of, at least, moderate physical activity a day [39] or accumulating <8 hours of sitting time a day [40]). Second, time spent on the website and clicks on optional pages were calculated. Optional pages included the quiz, page about social support, and the extra tips describing techniques such as "prompting rewards," "prompting a focus on past success," "providing instructions," "teaching to use prompts/cues," and "prompting self-talk." In addition, the average score on the quiz was calculated. Third, users' action plans were checked by CVdM for achievability and instrumentality toward the chosen behavior [41,42]. Fourth, we calculated how many users were able to (partially) reach their goals and how many times the goals were adapted. Finally, barrier identification/problem solving was checked for achievability (ie, is it possible to execute the solution?) and instrumentality (ie, does the solution actually solve the identified problem?) by CVdM; for example, the solution "scheduling a moment in my diary" was coded as instrumental and achievable for the problem "I do not have enough time," whereas this solution was considered achievable but not instrumental for the problem "I do not like to do it."

Figure 1. Flowchart of the first session.



Interviews were transcribed verbatim, and a directed content analysis was performed using NVivo Software (QSR International, Melbourne, Australia, Version 11, 2015) [43]. Content analysis is a way to comprise text into categories based on explicit coding rules [44-47]. In the directed content analysis, theory or prior research guides the coding. Directed content analysis is different from other strategies to analyze qualitative data in which codes most often emerge from the data [48]. Directed content analysis was considered best suited for our purpose because our coding scheme was based on previous research with “MyPlan 1.0” [30], and we were particularly interested in how participants precisely experienced the practical application of self-regulation techniques. Nevertheless, when a text fragment of the interview did not fit any of the predefined categories, a new category was created. Themes that did not contain enough data were not withheld. Coding was performed independently by two researchers (CVdM and LP). Furthermore, a weighted kappa was calculated, and it showed fair to good interrater agreement (weighted kappa=.67). Multimedia Appendix 3 shows an overview of the themes and subthemes. Multimedia Appendix 4 contains the completed COnsolidated criteria for REporting Qualitative research checklist [49].

Figure 2. Flowchart of the follow-up sessions.



Results

Demographic Characteristics

When contacted via telephone, 30 participants were willing to participate. However, 6 participants dropped out before the intervention period, and 4 participants did not respond to the researchers' telephone calls. Recruitment was continued until 20 participants fully completed the 5 intervention sessions. Table 1 shows the demographic characteristics of 20 participants.

Users' Website Data

Table 2 shows the users' website data according to both behaviors separately. Most users also visited the free-choice components, such as the quiz and additional pages regarding the social support. Only a small number of users indicated that they did not reach their goal or did not want to adapt their goal during the follow-up sessions. Almost all plans (38 of 40) were achievable and instrumental (eg, "On Monday and Wednesday I will perform my workout schedule at home"). The 2 exceptions were plans about sedentary behavior. In these plans, users indicated that they would perform a physical activity-related activity. During the first session, users had to select barriers and solutions from a list, which made all coping plans instrumental.

Interviews

Website in General

In general, users stated that participating in the study and being involved in the intervention program raised awareness of their own behavior.

> *You are also made more aware, and that's where it all starts.* [Woman, >45 years, high educational level, normal weight]
>
> *It was just the fact that I was more aware because I had to take a moment for it.* [Woman, 18-45 years, high educational level, normal weight]

Overall, the intervention website was perceived as user friendly and easy in use. Users highlighted the fact that it was clear and straightforward. In addition, the layout of the website was experienced as positive; it was simple and clear. Yet, some users would have liked a more colorful design.

Table 1. Demographic characteristics of participants.

Characteristics	Particpants (N=20)
Gender, n (%)	
Men	10 (50)
Women	10 (50)
Age in years, mean (SD), range	46.65 (16.65), 21-74
18-45 y, n (%)	10 (50)
>45 y, n (%)	10 (50)
Educational level, n (%)	
Primary education	1 (5)
Lower secondary education	1 (5)
Higher secondary education	8 (40)
College or university	10 (50)
Body mass index (kg/m²), mean (SD), range	25.42 (4.99), 18.47-37.81
Not overweight, n (%)	11 (55)
Overweight, n (%)	9 (45)

Table 2. Users' website data according to the 2 target behaviors (sedentary behavior and physical activity).

Website Data	Total (N=20)	Sedentary behavior (n=8)	Physical activity (n=12)
Number of users receiving feedback of not reaching the health norm, n (%)	7 (35)	6 (75)	1 (8)
Time spent per session (min)	6.67	6.74	6.59
Number of users reading the extra tips, n (%)	16 (80)	6 (75)	10 (83)
Number of users reading more about social support, n (%)	14 (70)	4 (50)	10 (83)
Number of users taking the quiz, n (%)	20 (100)	8 (100)	12 (100)
Mean score on the quiz (out of 5)	4.4	4.71	4.08
Number of users willing to monitor their behavioral change, n (%)	15 (75)	4 (50)	11 (92)
Number of plans not achievable or instrumental, n (%)	2 (2)	2 (2)	0 (0)
Indication of...during goal review, n (%)			
Total achievement	39 (49)	20 (63)	19 (40)
Partial achievement	37 (46)	11 (34)	26 (54)
Failure	4 (5)	1 (3)	3 (6)
Choice to...their plan, n (%)			
Adapt	11 (14)	5 (16)	6 (12)
Maintain	69 (86)	27 (84)	42 (88)
Number of solutions not achievable or instrumental, n (%)			
Session 2	8 (40)	2 (25)	6 (50)
Session 3	4 (20)	2 (25)	2 (17)
Session 4	7 (35)	1 (13)	6 (50)
Session 5	4 (20)	2 (25)	2 (17)

> *I thought it was a very good website. Very clear. I always knew what to do, where to click.* [Woman, 18-45 years, high educational level, normal weight]

> *It provided overview and was very clear. Nothing negative to mention. It was very easy, very simple. Yes, you could not do anything wrong I think.* [Man, >45 years, high educational level, normal weight]

> *I thought the layout was simple, but that didn't bother me. I think it contributed to the clarity.* [Man, 18-45 years, high educational level, normal weight]

In line with that, users also stated that they would have liked more interaction on the website and more new content per session. For some users, the website was too repetitive and could have been more appealing. Yet, most of the participants were positive about the website and the initiative in general.

> *I think, if people will visit the website regularly, they will want to see something new every time though.* [Woman, >45 years, low educational level, overweight]
>
> *It is useful that you try to let people be physically active. You can think about it yourself, everything comes from you. There is no one telling you: 'You have to do this if that happens'. You give yourself feedback.* [Woman, 18-45 years, low educational level, overweight]

Almost all participants experienced the intervention as personally relevant and appropriate. However, the website seemed less fitting for persons who considered themselves as being physically active or for individuals with a lack of motivation.

> *It is developed generation-independently, from 7 until 77 in a manner of speaking.* [Man, >45 years, low educational level, overweight]
>
> *Normally, I am already physically active. In that way, the added value for me was minimal. Maybe the intervention is too restricted because it is assumed that people experience difficulties in being physically active.* [Man, >45 years, high educational level, normal weight]

Most users appreciated the time efficiency of the website. Some users would have liked a little more content and for other users, content could have been shown in even less internet pages.

> *That (cf. the length) was very reasonable. Certainly not too long. However, not too short either. I had expected a lot more questions and other things.* [Man, >45 years, high educational level, overweight]

In addition, the intervention was perceived as motivating and stimulating for behavioral change by most users. However, some users experienced problems putting their intention into action. Other users were not motivated enough to change their behavior.

> *It is stimulating to initiate behavior.* [Man, >45 years, low educational level, overweight]
>
> *The website totally helped me, because I wasn't exercising anymore at all and now I am exercising again. So it did work.* [Man, 18-45 years, low educational level, overweight]
>
> *It is a very good initiative, but it is still difficult to translate it into action and actually move more or sit less. It seems evident, but it is not.* [Man, 18-45 years, high educational level, normal weight]

Goal Setting

Users often mentioned that the difference between physical activity and sedentary behavior was not clear for them, which made the intervention more complex.

> *For me there was little difference. If you sit less, then you automatically move more, and if you move more, then you sit less. So I didn't think it was clear.* [Woman, >45 years, high educational level, normal weight]

Providing Information on the Consequences of the Behavior

All participants stated that being more physically active or less sedentary has benefits for both physical and mental health. Some participants believed in the benefits but indicated that they had not experienced the benefits because of the intervention.

> *I think it has an influence. I really believe it has, but I have not experienced it.* [Woman, 18-45 years, high educational level, normal weight]

Accordingly, most users indicated that they did not learn new things through the intervention. They already knew the consequences of their behavior. They only had to be reminded to do something about it.

> *Learned new things? No. But it gave new insights, you take a moment to think about it.* [Man, >45 years, high educational level, normal weight]

Providing Feedback on Performance

The tailored feedback was highly appreciated by users. They recommended such feedback as the first step toward behavioral change. According to users, the feedback was personally tailored and made them aware that they had to change their behavior. Some users found that the feedback stimulated them actually to alter their behavior. Other users did not remember the feedback from the first session.

> *It was good to know where you are because you really don't have a clue.* [Woman, 18-45 years, high educational level, overweight]
>
> *I thought it (cf. the feedback) was good. That way, you know where you are and where you can improve. And it is different for every person. So, it is more personal.* [Woman, 18-45 years, high educational level, normal weight]

Action Planning

Action planning was experienced as highly motivating. Users appreciated the fact that they could plan their personal goals in a structured way by questions. Many users indicated that they actually performed their goal as planned.

> *I think it is important to plan this. Because everyone is busy and otherwise there is always something else coming up. If you don't make it a goal or plan in your week, it will not occur or it will fade with time.* [Woman, >45 years, high educational level, normal weight]

> *So putting my mobile phone further away (cf. in order to decrease sedentary time) is something that I do now.* [Man, 18-45 years, low educational level, normal weight]

Some users reported problems with action planning. They thought it was difficult to plan behavioral change a week in advance, especially when they had changing work hours. Furthermore, they preferred planning using a calendar rather than by questions. Other users found it difficult to plan behavioral change because they lacked the knowledge and inspiration about what to do. They wanted ready-to-use activity programs.

> *If you know what you want to do, but you do not put the words into action, then you fill this in. However, if someone knows he wants to be more physically active, but hedoes not know how exactly, then I think he will ask himself: "What should I do now?"* [Man, 18-45 years, high educational level, normal weight]

> *At the beginning I found it difficult to set up goals for myself.* [Woman, 18-45 years, high educational level, normal weight]

Barrier Identification/Problem Solving

Most users found it a good idea to think about barriers in advance and try to find solutions. However, many indicated that it was difficult to anticipate what could go wrong and how to overcome problems. Users expected the website to provide more guidance for this component.

> *What I really appreciated, is the fact that you were obliged to write down at least one barrier and how to cope with it. I had to take a bit of time to think about it, but in the end I always found one. The barrier component is the most powerful of the intervention.* [Man, 18-45 years, high educational level, normal weight]

> *Sometimes it was difficult. Because experiencing barriers is not difficult, but finding solutions is not always easy. Most of the time, the same barriers arose.* [Woman, 18-45 years, high educational level, normal weight]

> *Barrier identification really was something else (cf. in comparison to action planning). You have to be able to think immediately about what hinders you. That was more difficult. And maybe there could have been more guidance from the website.* [Woman, >45 years, low educational level, overweight]

Prompting Self-Monitoring of Behavior

Many users misunderstood the purpose of self-monitoring and wrote down their plan in advance to remind them about it, but did not keep track of whether they executed the planned behaviors or not.

> *I always wrote it down in my diary, in color. That is definitely useful, otherwise you forget about it.* [Man, 18-45 years, high educational level, normal weight]

> *I had expected that I would be assisted to monitor my goals myself, to see how my sitting time changes. But I was not asked to write down my sitting time.* [Man, 18-45 years, high educational level, normal weight]

Plan Social Support

There were a few users who commented on the social support component. Some users found it very useful to involve others, whereas other users preferred to keep their behavioral change more private.

> *I also appreciated the more practical tips such as inviting neighbors or not exercising alone. I found it nice to read and I often took it into account.* [Women, 18-45 years, high educational level, normal weight]

> *I did not really like the social parts. I prefer to do this on my own.* [Woman, >45 years, low educational level, normal weight]

Prompt Review of Behavioral Goals

The largest group of users found it useful to review their goals. Many users indicated that having to log in again was the most motivating part of the intervention.

> *The good thing was that it repeated itself every week. Another program ends after one session and then you have the tendency to put it aside. Since you had to log back in for five weeks, you wanted to do what they asked because they would ask if you did it.* [Woman, >45 years, low educational level, overweight]

Tips

Most users expressed their interest in the extra tips and found them very useful. The tips were experienced as feasible and inspiring. Especially, the tip regarding "using prompts or cues" was often implemented. Some users indicated that more new tips during the sessions were needed. Reading success stories of other possible users was also perceived as of added value to the website, although some stated that the stories were too predictable.

> *The tips were very interesting because they were practically feasible. It were simple tips that were achievable."* [Woman, >45 years, high educational level, normal weight]

> *It is always motivating to see (cf. read) how someone else does it, then you also want to motivate yourself to do it.* [Woman, >45 years, low educational level, overweight]

> *The most helping was the note on the fridge. It made you aware to not forget about your plans that day.* [Man, >45 years, low educational level, normal weight]

Discussion

Web-based interventions are increasingly used to alter health behaviors [10] and have shown to be more effective when grounded in a solid behavioral change theory [11]. However, the high levels of attrition highlight the importance to also target user engagement [36]. User engagement has been defined and measured in many ways [50]. According to Perski et al., engagement with a Web-based intervention is influenced by context (eg, the demographic characteristics of the population) and intervention (eg, the complexity of the intervention) variables [51]. This study focuses on the latter by investigating how users experienced a self-regulation-based eHealth intervention targeting physical activity and sedentary behavior. Users' website data were analyzed, and 20 semistructured interviews were performed.

Besides investigating users' opinions about self-regulation techniques, we also explored how they perceived the intervention in general. In comparison with the users of "MyPlan 1.0" [30], those of the 2.0-version appreciated the time efficiency and user-friendliness of the program; this is encouraging because it proves that an iterative approach in which users are consulted during the development of the intervention pays off [26]. Intervention developers should keep an eye on the user-friendliness of their intervention. We found that a simple but agreeable layout enhanced user-friendliness. Likewise, previous research has indicated that professional design and simple navigation can increase engagement [52]. Some users suggested that the development of a similar mobile app may further increase user-friendliness and interactivity; this suggestion is in line with research showing that the use of mobile apps might increase the adherence [53]. Most users perceived the sessions' duration of approximately 5 minutes as a perfectly reasonable length. Intervention developers are already aware that eHealth interventions should be kept short and to the point [32,52] but reducing length while still implementing different self-regulation techniques has not been an easy endeavor.

This study revealed that most users were well aware of the benefits of increasing physical activity or reducing sedentary behavior; this was reported in the interviews. Users often mentioned that the intervention did not substantially increase their knowledge about the beneficial effect of a more active lifestyle, and this finding was corroborated by the high scores on the quiz, which aimed to provide information engagingly. Notwithstanding, users were interested in information and all completed the optional quiz. The findings indicate that further tailoring and offering more advanced information is recommended in this target population. In addition, previous research highlights the importance of providing new information tailored to the users' needs [32]; for example, Short et al. stated that offering personalized information could increase men's engagement in a Web-based intervention targeting physical activity and nutrition [54]. Of further interest, reading information and receiving personal feedback on the questionnaires seemed to function as a prompt to behavioral change; it reminded users about the importance of adopting a more active way of living.

Of particular interest to this study were the experiences and opinions of users about the self-regulatory strategies to bridge the intention-behavior gap. Key to our eHealth intervention were action planning and problem solving. Action planning consisted of formulating specific actions and planning about when and how they will conduct these behaviors. Action planning seemed to be feasible. Few users stated unachievable plans and many were able to reach their goals, at least, partially. However, thinking in advance about actions was experienced as difficult and effortful by users. Some stated that it was difficult to come up with specific actions or plan these actions a week in advance, and this is a good remark. An improvement may be to allow users to create and evaluate specific plans on a daily basis. Implementing such microcycles might offer users more guidance in creating instrumental and achievable plans on a daily basis.

The implementation of the technique "barrier identification/problem solving" was less feasible. Many users struggled with identifying barriers and finding solutions in advance, especially in the follow-up sessions in which they had to answer an open-ended question; this was communicated in the interviews and further corroborated by the analysis of the provided barriers and solutions at the website. Our results seem to be at odds with those of other studies. Sniehotta et al. [55] successfully implemented this technique in their intervention to increase physical activity in cardiac rehabilitation patients; their implementation of the technique was very similar to ours—participants were asked which barriers could interfere with their plans and how they could successfully cope with these barriers. However, an important difference with our study is that trained consultants helped users with problem solving in face-to-face contact. Indeed, self-regulation techniques have mostly been used in face-to-face settings [13]. It may well be that counselors are better able to adapt to the implementation of these techniques to the context and needs of an individual. To date, Web-based interventions do not easily offer such an opportunity, and this is an issue worth further consideration and follow-up. Effective techniques may become useless (or even counterproductive) when their implementation is or remains suboptimal. Based on these findings, we recommend offering sufficient guidance when implementing the "barrier identification/problem solving" technique; for example, a button saying "need help?" was added in "MyPlan 2.0" When clicking on this button, users are shown an extensive list of potential barriers and solutions, which can guide them to answer the open-ended question.

In the interviews, some participants mentioned that the intervention may be of lesser use for individuals who are not ready for change yet, and this view is in line with various theoretical models of behavioral change, such as the Stages of Change Theory [56] and the Health Action Process Approach [57]. According to these models, individuals who are not ready to change will not engage in action programs. Indeed, studies investigating engagement according to user characteristics show that users' level of motivation is an important factor for the eHealth uptake [58]. Interventions targeting these individuals might then better include techniques such as motivational interviewing [59], focusing on raising awareness, and eliciting change talk. Such motivational techniques were largely absent in our intervention. We reasoned that eHealth interventions were relatively inadequate for participants with low motivation to change behavior in the short term. Perhaps, more intensive interventions, including face-to-face contact, may be more suited for these individuals [60].

In addition, users indicated that the intervention might be of lesser use for individuals who already have a habit of being active. Inadvertently, many of our participants already had an active way of living. Their personal feedback on the questionnaire stated that they reached the health norm. We had opted not to exclude participants who reached the health norms. First, research has demonstrated that individuals often overestimate their activity levels when self-report measures of physical activity are used [61]. We reasoned that participants may become more accurate of their estimations of physical activity by engaging in the intervention. Second, we reasoned that the eHealth intervention might also help in maintaining the behavior of those who are already habitually active, but this might not be the case. These individuals may experience the action and coping plan as needlessly effortful, frustrating, and cumbersome. Consequently, informing users explicitly about the target group of the intervention might be worth considering.

One of the strengths of this study was the diversity of the sample with an equal distribution of gender, age, educational level, and body mass index. Furthermore, having both users' website data, as well as interview data, strengthened our conclusions. Finally, the perspective of users on the specific implementation of self-regulation techniques has not been often investigated. The most important limitation of this study was the fact that we did not investigate the participants' actual levels of physical activity and sedentary behavior using validated methods. Consequently, we do not know whether our sample was more active than the general population. In addition, we were unable to assess the experiences of 4 users who quit the intervention. It may well be that their experience with the intervention was less positive. Furthermore, participants who were acquaintances of researchers might have had a more positive perception of the eHealth intervention. However, to limit this impact, these participants were always interviewed by a trained researcher they did not know.

In conclusion, this study reveals that behavioral change theories may be necessary but not sufficient to guarantee the efficacy in designing interventions. Equally important is the involvement of end users [25-27] because they can inform intervention developers on how self-regulation techniques should (or should not) be integrated. To ameliorate users' engagement with a Web-based intervention, we have the following recommendations: create short (5-6 minutes) interventions with a straightforward layout; provide novel and tailored information regarding the benefits of the health behavior; make users create specific action plans and review these plans in the follow-up sessions; and provide guidance and practical examples when adding a problem solving module.

Acknowledgments

The authors would like to thank Professor Dr Armand De Clercq for his support in developing "MyPlan 2.0."

Conflicts of Interest

None declared.

Multimedia Appendix 1

Implementation of the behavior change techniques in the website.

[PDF File (Adobe PDF File), 599KB - jmir_v20i10e10412_app1.pdf]

Multimedia Appendix 2

The Interview Guide.

[PDF File (Adobe PDF File), 39KB - jmir_v20i10e10412_app2.pdf]

Multimedia Appendix 3

Overview of the themes and subthemes.

[PDF File (Adobe PDF File), 28KB - jmir_v20i10e10412_app3.pdf]

Multimedia Appendix 4

Completed COREQ checklist.

[PDF File (Adobe PDF File), 45KB - jmir_v20i10e10412_app4.pdf]

References

1. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth?: a systematic review of published definitions. World Hosp Health Serv 2005;41(1):32-40. [Medline: 15881824]
2. Krebs P, Prochaska JO, Rossi JS. A meta-analysis of computer-tailored interventions for health behavior change. Prev Med 2010;51(3-4):214-221 [FREE Full text] [doi: 10.1016/j.ypmed.2010.06.004] [Medline: 20558196]
3. Ekeland AG, Bowes A, Flottorp S. Effectiveness of telemedicine: a systematic review of reviews. Int J Med Inform 2010 Nov;79(11):736-771. [doi: 10.1016/j.ijmedinf.2010.08.006] [Medline: 20884286]

4. Neuhauser L, Kreps GL. Rethinking Communication in the E-health Era. J Health Psychol 2003 Jan;8(1):7-23. [doi: 10.1177/1359105303008001426] [Medline: 22113897]
5. Broekhuizen K, Kroeze W, van Poppel MN, Oenema A, Brug J. A systematic review of randomized controlled trials on the effectiveness of computer-tailored physical activity and dietary behavior promotion programs: an update. Ann Behav Med 2012 Oct;44(2):259-286 [FREE Full text] [doi: 10.1007/s12160-012-9384-3] [Medline: 22767052]
6. de Vries H, Brug J. Computer-tailored interventions motivating people to adopt health promoting behaviours: introduction to a new approach. Patient Educ Couns 1999 Feb;36(2):99-105. [Medline: 10223015]
7. Kreuter MW, Wray RJ. Tailored and targeted health communication: strategies for enhancing information relevance. Am J Health Behav 2003 Dec;27 Suppl 3:S227-S232. [Medline: 14672383]
8. Kreuter M, Farrell D, Brennan L, Olevitch L. Tailoring Health Messages: Customizing Communication with Computer Technology. New York: Taylor & Francis, Inc; 1999.
9. Rimer BK, Kreuter MW. Advancing Tailored Health Communication: A Persuasion and Message Effects Perspective. J Communication 2006 Aug;56(s1):S184-S201. [doi: 10.1111/j.1460-2466.2006.00289.x]
10. Vandelanotte C, Müller AM, Short CE, Hingle M, Nathan N, Williams SL, et al. Past, Present, and Future of eHealth and mHealth Research to Improve Physical Activity and Dietary Behaviors. J Nutr Educ Behav 2016 Mar;48(3):219-228.e1. [doi: 10.1016/j.jneb.2015.12.006] [Medline: 26965100]
11. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
12. Sniehotta FF, Presseau J, Araújo-Soares V. Time to retire the theory of planned behaviour. Health Psychol Rev 2014;8(1):1-7. [doi: 10.1080/17437199.2013.869710] [Medline: 25053004]
13. Maes S, Karoly P. Self-Regulation Assessment and Intervention in Physical Health and Illness: A Review. Applied Psychology 2005 Apr;54(2):267-299. [doi: 10.1111/j.1464-0597.2005.00210.x]
14. Bandura A. Social cognitive theory of self-regulation. Organizational Behavior and Human Decision Processes 1991 Dec;50(2):248-287 [FREE Full text] [doi: 10.1016/0749-5978(91)90022-L]
15. Sniehotta F, Scholz U, Schwarzer R. Bridging the intention–behaviour gap: Planning, self-efficacy, and action control in the adoption and maintenance of physical exercise. Psychology & Health 2005 Apr;20(2):143-160 [FREE Full text] [doi: 10.1080/08870440512331317670]
16. 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]
17. 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]
18. Greaves CJ, Sheppard KE, Abraham C, Hardeman W, Roden M, Evans PH, et al. Systematic review of reviews of intervention components associated with increased effectiveness in dietary and physical activity interventions. BMC Public Health 2011;11:119 [FREE Full text] [doi: 10.1186/1471-2458-11-119] [Medline: 21333011]
19. Stadler G, Oettingen G, Gollwitzer PM. Physical activity in women: effects of a self-regulation intervention. Am J Prev Med 2009 Jan;36(1):29-34. [doi: 10.1016/j.amepre.2008.09.021] [Medline: 18977113]
20. Anderson ES, Wojcik JR, Winett RA, Williams DM. Social-cognitive determinants of physical activity: the influence of social support, self-efficacy, outcome expectations, and self-regulation among participants in a church-based health promotion study. Health Psychol 2006 Jul;25(4):510-520. [doi: 10.1037/0278-6133.25.4.510] [Medline: 16846326]
21. Plaete J, De Bourdeaudhuij I, Verloigne M, Crombez G. Acceptability, feasibility and effectiveness of an eHealth behaviour intervention using self-regulation: 'MyPlan'. Patient Educ Couns 2015 Jul 26. [doi: 10.1016/j.pec.2015.07.014] [Medline: 26277282]
22. 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]
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. Spittaels H, De Bourdeaudhuij I, Vandelanotte C. Evaluation of a website-delivered computer-tailored intervention for increasing physical activity in the general population. Prev Med 2007 Mar;44(3):209-217. [doi: 10.1016/j.ypmed.2006.11.010] [Medline: 17197015]
25. Whitehouse SR, Lam P, Balka E, McLellan S, Deevska M, Penn D, et al. Co-Creation With TickiT: Designing and Evaluating a Clinical eHealth Platform for Youth. JMIR Res Protoc 2013 Oct 18;2(2):e42 [FREE Full text] [doi: 10.2196/resprot.2865] [Medline: 24140595]
26. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]

27. De Vito Dabbs A, Myers BA, Mc Curry KR, Dunbar-Jacob J, Hawkins RP, Begey A, et al. User-centered design and interactive health technologies for patients. Comput Inform Nurs 2009;27(3):175-183 [FREE Full text] [doi: 10.1097/NCN.0b013e31819f7c7c] [Medline: 19411947]
28. Bardus M, Blake H, Lloyd S, Suzanne Suggs L. Reasons for participating and not participating in a e-health workplace physical activity intervention. Intl J of Workplace Health Mgt 2014 Nov 04;7(4):229-246 [FREE Full text] [doi: 10.1108/IJWHM-11-2013-0040]
29. Brouwer W, Oenema A, Crutzen R, de Nooijer J, de Vries NK, Brug J. What makes people decide to visit and use an Internet-delivered behavior-change intervention? A qualitative study among adults. Health Education 2009:460-473 [FREE Full text] [doi: 10.1108/09654280911001149]
30. Poppe L, Van der Mispel C, De Bourdeaudhuij I, Verloigne M, Shadid S, Crombez G. Users' thoughts and opinions about a self-regulation-based eHealth intervention targeting physical activity and the intake of fruit and vegetables: A qualitative study. PLoS One 2017;12(12):e0190020 [FREE Full text] [doi: 10.1371/journal.pone.0190020] [Medline: 29267396]
31. Donkin L, Glozier N. Motivators and motivations to persist with online psychological interventions: a qualitative study of treatment completers. J Med Internet Res 2012;14(3):e91 [FREE Full text] [doi: 10.2196/jmir.2100] [Medline: 22743581]
32. Yardley L, Morrison LG, Andreou P, Joseph J, Little P. Understanding reactions to an internet-delivered health-care intervention: accommodating user preferences for information provision. BMC Med Inform Decis Mak 2010 Sep 17;10:52 [FREE Full text] [doi: 10.1186/1472-6947-10-52] [Medline: 20849599]
33. Plaete J, Crombez G, Van der Mispel C, Verloigne M, Van Stappen V, De Bourdeaudhuij I. Effect of the Web-Based Intervention MyPlan 1.0 on Self-Reported Fruit and Vegetable Intake in Adults Who Visit General Practice: A Quasi-Experimental Trial. J Med Internet Res 2016 Feb 29;18(2):e47 [FREE Full text] [doi: 10.2196/jmir.5252] [Medline: 26929095]
34. Van Dyck D, Plaete J, Cardon G, Crombez G, De Bourdeaudhuij I. Effectiveness of the self-regulation eHealth intervention 'MyPlan1.0.' on physical activity levels of recently retired Belgian adults: a randomized controlled trial. Health Educ Res 2016 Oct;31(5):653-664. [doi: 10.1093/her/cyw036] [Medline: 27422898]
35. Degroote L, Plaete J, De Bourdeaudhuij I, Verloigne M, Van Stappen V, De Meester A, et al. The Effect of the eHealth Intervention 'MyPlan 1.0' on Physical Activity in Adults Who Visit General Practice: A Quasi-Experimental Trial. Int J Environ Res Public Health 2018 Jan 30;15(2) [FREE Full text] [doi: 10.3390/ijerph15020228] [Medline: 29385770]
36. Van der Mispel C, Poppe L, Crombez G, Verloigne M, De Bourdeaudhuij I. A Self-Regulation-Based eHealth Intervention to Promote a Healthy Lifestyle: Investigating User and Website Characteristics Related to Attrition. J Med Internet Res 2017 Jul 11;19(7):e241 [FREE Full text] [doi: 10.2196/jmir.7277] [Medline: 28698168]
37. Lee PH, Macfarlane DJ, Lam TH, Stewart SM. Validity of the International Physical Activity Questionnaire Short Form (IPAQ-SF): a systematic review. Int J Behav Nutr Phys Act 2011;8:115 [FREE Full text] [doi: 10.1186/1479-5868-8-115] [Medline: 22018588]
38. Wijndaele K, De Bourdeaudhuij I, Godino JG, Lynch BM, Griffin SJ, Westgate K, et al. Reliability and validity of a domain-specific last 7-d sedentary time questionnaire. Med Sci Sports Exerc 2014 Jun;46(6):1248-1260 [FREE Full text] [doi: 10.1249/MSS.0000000000000214] [Medline: 24492633]
39. Haskell WL, Lee I, Pate RR, Powell KE, Blair SN, Franklin BA, et al. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. Med Sci Sports Exerc 2007 Aug;39(8):1423-1434. [doi: 10.1249/mss.0b013e3180616b27] [Medline: 17762377]
40. van der Ploeg HP, Chey T, Korda RJ, Banks E, Bauman A. Sitting time and all-cause mortality risk in 222 497 Australian adults. Arch Intern Med 2012 Mar 26;172(6):494-500. [doi: 10.1001/archinternmed.2011.2174] [Medline: 22450936]
41. Plaete J, De Bourdeaudhuij I, Verloigne M, Crombez G. The use and evaluation of self-regulation techniques can predict health goal attainment in adults: an explorative study. PeerJ 2016;4:e1666 [FREE Full text] [doi: 10.7717/peerj.1666] [Medline: 26966648]
42. van Osch L, Lechner L, Reubsaet A, De Vries H. From theory to practice: An explorative study into the instrumentality and specificity of implementation intentions. Psychol Health 2010 Mar;25(3):351-364. [doi: 10.1080/08870440802642155] [Medline: 20204939]
43. NVivo. Qualitative Data Analysis Software. 11 ed: QSR International Pty Ltd; 2015.
44. Stemler S. An overview of content analysis. Practical assessment, research & evaluation 2001;7(17):1531-7714.
45. Albig W. BERELSON, BERNARD. Content Analysis in Communication Research. Pp. 220. Glencoe, Ill.: The Free Press, 1952. $3.50. The ANNALS of the American Academy of Political and Social Science 2016 Sep 08;283(1):197-198 [FREE Full text] [doi: 10.1177/000271625228300135]
46. Krippendorff K. Content Analysis: An Introduction to Its Methodology. California: SAGE Publications; 2012.
47. Weber R. Basic Content Analysis. California: SAGE Publications; 1990.
48. 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]
49. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]

50. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and Promoting Effective Engagement With Digital Behavior Change Interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
51. 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 2016 Dec 13. [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
52. Brouwer W, Oenema A, Crutzen R, de Nooijer J, de Vries N, Brug J. What makes people decide to visit and use an internet‐delivered behavior‐change intervention? Health Education 2009 Oct 16;109(6):460-473 [FREE Full text] [doi: 10.1108/09654280911001149]
53. Carter MC, Burley VJ, Nykjaer C, Cade JE. Adherence to a smartphone application for weight loss compared to website and paper diary: pilot randomized controlled trial. J Med Internet Res 2013;15(4):e32 [FREE Full text] [doi: 10.2196/jmir.2283] [Medline: 23587561]
54. Short CE, Vandelanotte C, Dixon MW, Rosenkranz R, Caperchione C, Hooker C, et al. Examining participant engagement in an information technology-based physical activity and nutrition intervention for men: the manup randomized controlled trial. JMIR Res Protoc 2014;3(1):e2 [FREE Full text] [doi: 10.2196/resprot.2776] [Medline: 24389361]
55. Sniehotta FF, Scholz U, Schwarzer R. Action plans and coping plans for physical exercise: A longitudinal intervention study in cardiac rehabilitation. Br J Health Psychol 2006 Feb;11(Pt 1):23-37. [doi: 10.1348/135910705X43804] [Medline: 16480553]
56. Prochaska J, DiClemente C. Toward a comprehensive model of change. In: Treating Addictive Behaviors. Berlin: Springer; 1986:3-27.
57. 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]
58. Hardiker NR, Grant MJ. Factors that influence public engagement with eHealth: A literature review. Int J Med Inform 2011 Jan;80(1):1-12. [doi: 10.1016/j.ijmedinf.2010.10.017] [Medline: 21112244]
59. Rollnick S, Miller W. What is Motivational Interviewing? Behav. Cognit. Psychother 2009 Jun 16;23(04):325-334 [FREE Full text] [doi: 10.1017/S135246580001643X]
60. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
61. Rzewnicki R, Vanden AY, De Bourdeaudhuij I. Addressing overreporting on the International Physical Activity Questionnaire (IPAQ) telephone survey with a population sample. Public Health Nutr 2003 May;6(3):299-305. [doi: 10.1079/PHN2002427] [Medline: 12740079]

Abbreviations

IPAQ: International Physical Activity Questionnaire

Edited by G Eysenbach; submitted 14.03.18; peer-reviewed by P Krebs, A Maier; comments to author 19.04.18; revised version received 05.06.18; accepted 28.06.18; published 01.10.18.

Please cite as:
Poppe L, Van der Mispel C, Crombez G, De Bourdeaudhuij I, Schroé H, Verloigne M
How Users Experience and Use an eHealth Intervention Based on Self-Regulation: Mixed-Methods Study
J Med Internet Res 2018;20(10):e10412
URL: https://www.jmir.org/2018/10/e10412/
doi:10.2196/10412
PMID:30274961

Review

Clinical Videoconferencing as eHealth: A Critical-Realist Review and Qualitative Meta-Synthesis

Anne Granstrøm Ekeland[1,2], Dr rer pol; Anne Helen Hansen[3,4], MD, PhD; Trine Strand Bergmo[1], PhD

[1]Norwegian Centre for eHealth Research, University Hospital of North Norway, Tromsø, Norway
[2]Department of Clinical Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway
[3]Centre for Quality Improvement and Development, University Hospital of North Norway, Tromsø, Norway
[4]Department of Community Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway

Corresponding Author:
Anne Granstrøm Ekeland, Dr rer pol
Norwegian Centre for eHealth Research
University Hospital of North Norway
PO Box 35
University Hospital of North Norway
Tromsø, 9038
Norway
Phone: 47 95266791
Email: anne.granstrom.ekeland@ehealthresearch.no

Abstract

Background: Earlier work has described videoconferencing technologies, peripheral equipment, organizational models, human relations, purposes, goals and roles as versatile, multifaceted, and those used differently in different clinical practices. Knowledge about benefits and challenges connected to specific characteristics of services are lacking. A 2005 systematic review of published definitions of electronic health (eHealth) identified 51 unique definitions. In addition, the “10 E’s of eHealth” was developed. In 2015, the question “What Is eHealth: Time for an Update?” was posed.

Objective: Considering videoconferencing as eHealth, the objective of the paper is twofold: to demonstrate and cluster (different) clinical videoconferencing practices and their situated implications and to suggest interpretive concepts that apply to all clusters and contribute to generative learning of eHealth by discussing the concepts as add-ons to existing descriptions of eHealth in the “10 E’s of eHealth.”

Methods: We performed a literature search via the National Center for Biotechnology Information, encompassing PubMed and PubMedCentral, for quality reviews and primary studies. We used the terms “videoconferencing” and “clinical practices.” The selection process was based upon clearly defined criteria. We used an electronic form to extract data. The analysis was inspired by critical and realist review types, grounded theory, and qualitative meta-synthesis.

Results: The search returned 354 reviews and primary studies. This paper considered the primary studies, and 16 were included. We identified the following 4 broad clusters: videoconferencing as a controlled technological intervention within existing health care organizations for expert advice, controlled mixed interventions with experimental organizational arrangements, videoconferencing as an emerging technosocial service involving dialogue and empowerment of patients, and videoconferencing as a controlled intervention to improve administrative efficiency. The analysis across the clusters resulted in a proposal to add the following 4 D’s to the existing 10 E’s: (inter)-dependent, differentiated across services and along temporal lines, dynamic in terms of including novel elements for meeting incremental needs, and demanding in terms of making new challenges and dual results visible and needing fresh resources to meet those challenges. For a normative discussion about what eHealth should be according to authors’ conclusions, results suggested ethical, in that users interests should be respected, and not harmful in terms of increasing symptom burden.

Conclusions: Services were enacted as dynamic, differentiated concerning content and considerations of quality and adaptive along temporal lines. They were made to work from an ongoing demand for fresh resources, making them interdependent. The 4 D’s—Dynamic, Differentiated, Demanding, and (inter) Dependent—serve as pragmatic add-ons to the “10 E’s of eHealth.” Questions concerning outcome of specified balances between standardization and customization in clinical settings should be addressed in future research along with the emerging dual character of outcome: services being considered both “good” and “bad.”

(J Med Internet Res 2018;20(10):e282) doi:10.2196/jmir.8497

KEYWORDS

videoconferencing; clinical practice patterns; realist review; situated implications; eHealth; telemedicine

Introduction

Background

Videoconferencing has been used for half a century in health care. At its most sophisticated, it provides transmission of full-motion video and high-quality audio among multiple locations [1]. An early description of clinical videoconferencing involved real-time visits through cameras and televisions [2]. Such visits are still in clinical use and may include additional peripheral equipment such as stethoscopes, otoscopes, and derma scopes.

In 2010, Whitten et al [3] summarized the most popular areas for telemedicine through videoconferencing to be telepsychiatry [4], pediatric emergency consults [5], stroke diagnosis and treatment [6], teledialysis [7], and teledermatology [8].

Videoconferencing is part of electronic health (eHealth), and definitions and compositions of such services have developed along with technological refinement, changes in roles and responsibilities in health care and other contextual conditions. A 2005 systematic review identified 51 unique definitions of eHealth and concluded that it had become an accepted neologism despite the lack of clear, precise, agreed-upon definitions [9].

Challenges regarding the understanding of the notion, of use and nonuse, and the quality of such services have frequently been reported and can be connected with problems of determining what the intervention exactly consists of, as well as generalizable causes and effects due to the complexity of contexts [10]. In addition, earlier papers and reviews have demonstrated that videoconferencing often combines with customized organizational models and routines and specified human resources ranging from patients to family caregivers, nurses, and doctors, inhibiting different roles [11].

Moreover, complexities are demonstrated in studies considering videoconferencing as a component of interdependent relations among different technologies, the complexity of health care and the rituals and habits of patients and other stakeholders [12]. Furthermore, different purposes of videoconferencing have been described. In a 2014 systematic review of clinical use, 91% of included papers reported different clinical purposes, including diagnosis, treatment, counseling, and monitoring in a wide range of disciplines and settings [13].

The accounts above point to a complex field and the time seems to be right for obtaining contextualized knowledge of what kinds of services that work, for whom and under which circumstances. Granja et al also pointed this out in their recent systematic review of factors determining the success and failure of eHealth interventions [14]. They concluded that quality of care was most frequently mentioned as contributing to the success, and costs most frequently mentioned as contributing to the failure. In addition, they pointed to a critical need to perform in-depth studies of the workflow(s) that eHealth interventions support and to perceive the clinical processes involved.

This paper assumes that videoconferencing is interdependent with contexts and that workflow and quality consequently will be enacted and considered differently in different clinical processes. The paper provides an in-depth account of videoconferencing services, user patterns, and considerations of quality in different clinical practices.

The account is based on results of a critical-realist literature review, performed to explore what kinds of videoconferencing services were associated with what kinds of quality in different clinical processes, as presented in the academic literature [15].

We first present the results as a thick description of user patterns, by developing a narrative synthesis of comparable and contrasting services and considerations of quality. This synthesis is presented in clusters empirically ordered by similar technosocial configurations. Notably, this approach is grounded.

Second, we interpret the findings through an in-depth analysis [16]. We performed an exercise of reciprocal translation of the clusters and suggested and argued for common concepts and metaphors that are played out and can be applied to all clusters. This is a constructive analytical technique described under the theoretical umbrella of qualitative meta-synthesis [17].

Finally, after this exercise, we move on to discuss the concepts related to previous characteristics of eHealth. For this purpose, we rely on definitions and the “10 E’s of eHealth” as developed by Eysenbach [18]. The objective is to use our concepts as input for generative learning and further discussion of the concept of eHealth and considerations of quality.

In the following section, we provide an account of the critical-realist review, grounded perspective, concept of enactment, and steps in the qualitative meta-analysis as examples of research to obtain generative learning for the pragmatic adaptation of concepts.

We hope to get a deeper insight into prerequisites and impact in different health care practices [19]. Similarly, we hope to contribute to the knowledge of complexity and variety and, consequently, different understandings of eHealth [20].

This review can also be considered as an attempt to respond to a recent call for “new standards of research quality, namely (for example) rich theorizing, generative learning, and pragmatic adaptation to changing contexts in open systems characterized by dynamically changing interrelationships and tensions” [21].

In 2001, Gunther Eysenbach defined eHealth as a concept as follows: “In a broader sense, the term characterizes not only a technical development but also a state-of-mind, a way of thinking, an attitude, and a commitment for networked, global thinking, to improve health care locally, regionally, and worldwide using information and communication technology.”

[18] This definition was also elaborated on through 10 characteristics, Eysenbach's 10 E’s of eHealth [18].

In a following series of publications, different characteristics were developed [22,23]. In 2015, the question “What Is eHealth: Time for an Update?” was posed [24].

Theoretical Approaches to the Study

The paper refers to 2 different types of literature review, namely realist review and critical review [15,25]. The objectives also point to the specified features of grounded theory [16]. In the methods section, we elaborate the interconnection with grounded theory and the analytical perspective of meta-synthesis.

Realist Review and the Account of Diversity

Through realist review, we addressed the complexity of videoconferencing. Traditional methods of review may focus on measuring and reporting on generalizable program effectiveness; these often find that the evidence is mixed or conflicting, and provide little or no help to understand the specified features of services that worked or did not work in different contexts for different stakeholders or different purposes [25].

Realist review is a model of research synthesis designed to work with complex social interventions or programs, based on the emerging “realist” approach to evaluation. It provides an explanatory analysis aimed at discerning what works (descriptions of compositions of technologies, human resources, relations, and organizational arrangements) for whom (actors and roles), in what circumstances (clinical setting and socioeconomic setting if described), and in what respects (outcome measures and considerations of quality). Furthermore, a realist review often assesses how services work in terms of technological performance; of note, this element is not discussed in detail in our review.

Critical Review and Conceptual Innovation

A critical review goes beyond mere description of identified services and includes analysis and conceptual innovation [15]. Grant and Booth [15] argued that its product most typically manifests in a hypothesis or a model, not a categorical answer. While such a review serves to aggregate the literature on a topic, the interpretative elements, as in developing concepts, are necessarily subjective and the resulting product is the starting point for further discussion, not an endpoint. We pursue the development of metaphors or concepts to capture novel features that apply to all empirical clusters.

How can we shed light on the concept of eHealth through interpretations and conceptual constructs that apply to or emerge within all clusters? The meta-synthesis method described in the Methods section provides the strategy for such analyses. In addition, a critical review may also attempt to resolve competing schools of thought. As such, it may provide a “launch pad” for a new phase of conceptual development and subsequent testing. We will not address competing schools of thought but add-ons to existing concepts of eHealth.

For this purpose, we use the findings to compare and contrast with previous theoretical standpoints concerning eHealth as described in the 10 E’s. This is a deductive turn, and we relate our inductively derived concepts to the existing theory of eHealth.

Methods

Study Design

On September 15, 2015, we performed a systematic search for primary studies of videoconferencing in clinical practice. We searched through the National Center for Biotechnology Information, encompassing PubMed and PubMedCentral, using the terms “videoconferencing” and “clinical practices.” In addition, we included additional references from suggestions on the National Center for Biotechnology Information Web page or reference lists of papers identified in the original searches. This search was updated on May 15, 2018.

Inclusion Criteria

The review included primary studies published in English between January 1, 2010, and May 15, 2018, which reportedly implemented clinical videoconferencing or Web conferences with the synchronous interaction between, at least, 2 geographical locations. Included in services were technological interventions, singular or as part of a composite service, and the paper should include a clear research objective or question with a clearly described method. Both qualitative and quantitative methods were included.

Exclusion Criteria

Videoconferencing for educational or guideline development purposes and psychiatric or psychological interventions were excluded if these were not combined with somatic purposes. The reasons for excluding psychiatry and psychology were pragmatic, as the search yielded a large number of papers addressing such interventions. In addition, we excluded monitoring of vital signs and self-help apps if not combined with synchronous video or Web communications. Furthermore, feasibility and pilot studies were excluded.

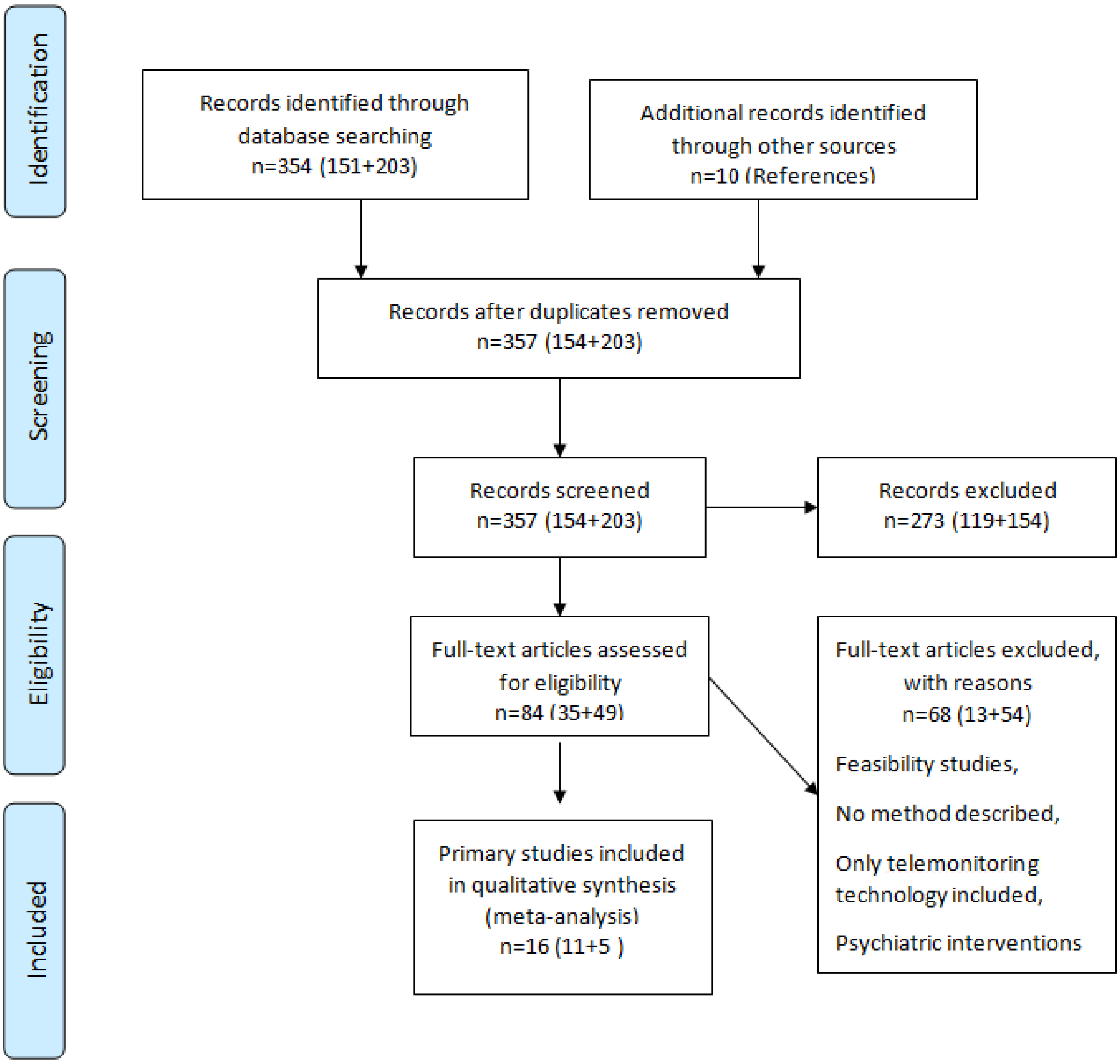
The Preferred Reporting Items for Systematic Reviews and Meta-Analyses Diagram

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram lists the steps of the inclusion and exclusion of papers (Figure 1).

Descriptive Summary

An electronic form was developed and used to extract data from the included papers (see Multimedia Appendix 1). These data items were selected as follows: first author and title, year published, population and clinical area, description of the intervention, additional technologies included in the intervention, participatory or organizational arrangements, outcomes, goals, conclusion, challenges, suggestions for service improvements, and suggestions for further research. Multimedia Appendix 2 provides a table of results, created using Excel.

Figure 1. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram.



Enactment of Clusters and Grounded Theory

The concept of enactment is used to suggest which practices of videoconferencing may play out differently through interdependencies [26,27]. A grounded perspective as described by Strauss and Corbin implies that empirical enactments should be the basis for deriving theoretical concepts, an inductive approach [16]. Thus, grounded theory is different from the deductive research model, that is, where the researcher chooses an existing theoretical framework (here the 10 E’s of eHealth) and collects data to show how the theory does or does not apply to the phenomenon under study.

We formed clusters by comparing and contrasting features responding to the realist review categories to specify the descriptive elements—what works and in what clinical setting. We coded for content, emerging themes, patterns, and novel ideas to also consider the 2 additional realist review themes, responding to what counts as good, for whom (actors and roles), and in what respects (outcome measures, considerations of quality, and suggestions for improvement). Furthermore, we used comparative methods involving open, axial, and selective coding of data where units of text (words, phrases, sentences, or paragraphs) were labeled, compared, and grouped [16].

Meta-Synthesis and the Development of Concepts

A conceptual meta-synthesis is an interpretive, analytical technique that uses qualitative findings reported in previous studies as building blocks for gaining a deeper understanding of particular phenomena [17].

To perform a meta-synthesis, we used the clustered summary to interpret and suggest concepts pertaining to the descriptions and considerations to construct conceptual generalizations. As data were rereviewed, we grouped categories that formed the basis for suggesting the theoretical concepts.

Our meta-synthesis attempted to integrate results from a number of different but interrelated qualitative and quantitative studies. The technique has an interpretive, rather than aggregating, intent, in contrast to the meta-analysis of quantitative studies.

Proposing New Concepts for Generative Learning: the “10 E’s of eHealth”

Within the clustered summaries, we first very briefly identified examples to provide connections with Eysenbach's 10 E’s.

Through this strategy, we showed how the services we investigated aligned with or accommodated the 10 E's. But the main focus of the analysis was to propose and argue for new concepts to contribute to generative learning, that is, what might be added to the 10 E's from our findings and interpretations? This implies that we suggest concepts for pragmatic adaptation to the existing 10 E's.

Results

Descriptive Summary

The 16 included primary studies covered 13 clinical areas, some of which were overlapping—chronic neurological care, dermatology and heart disease, stroke, neonatal care, primary health care, neuromuscular disease, pathology, palliative care, chronic care [hypertension, chronic obstructive pulmonary disease (COPD), and cardiovascular diseases combined], primary headache, pediatric telegenetics, rehabilitation of total knee replacement, and geriatric rehabilitation. In addition, we included outpatient triage. Studies reporting both quantitative and qualitative methods were included.

We have provided an overview of the data extracted from each of the included primary studies by listing the first author and title, year published, patient populations, clinical area, technosocial interventions, outcome measures, results, conclusion, any suggestions for improvement of the intervention and further research as an appendix (Multimedia Appendix 1).

We derived 4 broad clusters based on similarities in technosocial compositions. The clusters are as follows: videoconferencing as a controlled technological intervention within existing health care organizations for expert advice, including 5 papers [28-32]; controlled mixed interventions with experimental organizational arrangements, including 2 papers [33,34]; videoconferencing as an emerging technosocial service involving dialogue and empowerment of patients, including 8 papers [35-42]; and videoconferencing as a controlled intervention to improve the administrative efficiency, including 1 paper [43].

We suggested adding the 4 D's to Eysenbach's 10 E's of eHealth — eHealth is also (inter) dependent, differentiated across services and along temporal lines, dynamic in terms of incrementally including novel elements for meeting needs, and demanding in terms of making new challenges and dualistic results visible, and being in need of fresh resources to meet emerging challenges.

For a normative consideration, what eHealth also should be according to authors' conclusions, we added ethical, in that users' interests should be respected, and not harmful in terms of increasing the symptom burden.

Clusters, Characteristics, and Suggestions for Improvement

The 4 clusters were organized and elaborated according to the categories of realist review—what works (compositions of technologies, human resources, relations, and organizational arrangements), for whom (actors and roles), and in what circumstances (clinical setting and socioeconomic setting if described). This account makes up the descriptive part of the results, and we have clustered services that are comparable concerning the composition. We further considered the outcome and quality by assessing authors' proposals for improvement of services, to respond to the question of how services are considered and what counts as good (Multimedia Appendix 1).

Videoconferencing as a Controlled Expert Intervention Within Existing Organizational Arrangements

What Works, for Whom, in Which Clinical Area, and With Which Outcome: What eHealth Is

In chronic neurological care, videoconferencing was used in follow-ups from specialists to veterans living in rural areas, after an initial complete historical and neurological examination. Patients were highly satisfied with the outcomes of convenience and quality [28].

In dermatology and heart disease, specialist care was delivered in a primary health care center in rural areas. A majority of patients considered the services as a positive experience, particularly, as specialist visits were expensive [29]. In the UK National Health Service, a regional telestroke network comprising 7 district general hospitals delivered thrombolysis with effective management, access, and safety as an outcome; this was conditionally successful [30]. In medical treatment for a primary headache, the service consisted of face-to-face interviews, examination, cranial magnetic resonance imaging, and electroencephalogram, a discussion between 2 physicians at different locations, and prescription of medication, all through data and video transmission. The network demonstrated safe and efficacious delivery of services [31].

In weekly specialist palliative care, teleconsultations for patients with advanced cancer were compared with "care as usual" [32]. Furthermore, the patient-experienced symptom burden was discussed.

Considerations and Proposals for Improvement of Services' Quality

During assessments, complexities appeared that problematized generalizable conclusions for all services, and further success seemed to rely on the customization of certain aspects. For instance, results on the quality were narrowed to apply to more specified conditions in chronic neurological care [28], going through several encounters were considered a prerequisite to improving patients' experience of safety in dermatology and heart disease [29], and variability of success between local sites was reported in thrombolysis services [30]. In addition, further success turned out to depend on additional resources involving expertise from nurses or technicians [28] and social and human relations like collaboration and confidence [29]. Telemedicine did not necessarily lead to a better quality of advanced cancer care. Indeed, the use of telemedicine created a situation in which patients experienced a higher symptom burden, despite high degrees of satisfaction. Authors proposed further research on ways to optimize multidisciplinary care by teleconsultations and to decide appropriate timing and frequency of palliative care teleconsultations [32].

eHealth was enacted as conditionally successful and partly with contrasting results as in palliative care. eHealth was demanding new resources and customization of organizational, diagnostic, and ethical aspects. The balance between standardization or customization was played out as a challenge, as was the balance between fixed or dynamic and adaptive solutions.

Mixed Services With Novel Organizational Approaches for Expert Advice

What Works, for Whom, in Which Clinical Area, and With Which Outcome?

A novel approach in pediatric telegenetic services was described, comprising a geneticist, pediatrician, and genetic counselor team [33]. A telegenetics clinic offered a viable solution to providing competent and convenient access to a geneticist for patients in chronically underserved regions. For telerehabilitation after total knee replacement in Italy, a cost-effectiveness and cost-utility analysis was performed of a mixed telerehabilitation-standard rehabilitation program compared with usual care. Cost savings were documented [34].

Considerations and Proposals for Improvement of Services' Quality

Further studies were requested to define the outcomes of pediatric genetic evaluations better and specify which outcomes most appropriately helped to determine the satisfaction and efficacy of telegenetics evaluations compared with in-person genetic evaluations [33]. Uncertainty related to costs and long-term clinical outcomes raised important topics for future research in the rehabilitation study [34]. Moreover, in this cluster, more specified knowledge of outcomes, that is, the quest for more detailed and customized differentiation of services to understand quality was communicated. The dynamic adaptation of services was requested following the uncertainty of long-term effects.

Videoconferencing as Part of a Composite Service Involving Dialogue and Empowerment of Patients

What Works, for Whom, in Which Clinical Area, and With Which Outcome?

Birth delivery patterns in neonatal care in a state network was addressed in weekly educational videoconferences to establish guidelines for obstetrical, neonatal, and pediatric care at rural hospitals and telenursery sites [35]. In addition, the network maintained a 24-hour call center staffed by experienced nurses who provided case management for patients and their physicians across the state, including appropriate transfer of high-risk patients to regional perinatal centers associated with high-quality delivery. The service was considered successful on all outcome measures, particularly for rural and underserved populations.

In neuromuscular diseases, 2 different telecare protocols were followed depending on the severity of the patient's condition [36]. One group combined videoconferencing with face-to-face consultation once a month to see the patient in person and check the equipment, whereas the other maintained face-to-face consultation. The system was effective for home treatment and reduced the need for hospital admissions.

The Eastern Quebec Telepathology Network included a macroscopy station and 2 videoconferencing devices equipped with drawing tablets. The assessment considered the real-time evaluation of the concordance rate, turnaround time, and effects of telepathology on health care professionals, patients, organization, and delivery of care, from 3 years' experience [37]. The service reduced isolation and insecurity among pathologists working alone.

Interdisciplinary team group videoconference meetings every second week with the involvement of informal caregivers was as part of a total home-care program for hospice patients [38]. The role of technology was a mixed experience, and short interaction length was sometimes frustrating, but most caregivers reported feeling part of the team and were positive about the technology experience. Caregivers had positive relationships with hospice staff, felt involved in decision making, and got questions answered.

"Tile-Ippokratis," was an integrated platform providing low-cost eHealth services to elderly patients with chronic diseases and postsurgery patients with hypertension, COPD, and cardiovascular disease [39]. The platform reduced hospital admissions, provided safety, self-management, and quality of life. Users from both server and client units expressed positivity on the interface and data entering procedures of the devices used.

In outpatient palliative care, videoconferencing facilitated empathic patient–professional relationships [40]. Owing to rapid technological developments, the following 2 teleconsultation devices were used during the study: a desktop computer and an iPad 2. Results were focused attention and listening, the empathic engagement between patients and palliative care specialists, and an opportunity for patients to co-design their own care within the comfort of their homes. The potential of teleconsultations jeopardizing privacy was reported.

Another study addressed the same service with a focus on professional collaboration [41]. Synchronous audiovisual teleconsultation between a hospital-based specialist palliative care team and home-based palliative care patients was added to an existing community care model. The introduction of specialist teleconsultation in palliative home care supported multidisciplinary care.

For older rehabilitation, patients' telerehabilitation by weekly home visits was combined with monitoring of activity data and videoconference via iPad [42]. Positive outcomes were reported in terms of the experience, activity levels, fitness, functioning, and well-being.

Considerations and Proposals for Improvement of Services

In neonatal care, longitudinal knowledge on system-oriented changes was requested [35]. In neuromuscular diseases, adaptive organizational procedures and face-to-face consultations should be more developed to improve services [36]. In telepathology, a strong communication plan and highly coordinated efforts among surgeons, pathologists, stakeholders, laboratory staff, and biomedical, administrative, and information technology support teams working in different sites were considered crucial for further development [37]. In elderly care, palliative and hospice care, familiarization of technologies, more frequent meeting times, and additional training for family caregivers were needed [38,40]. In addition, further suggestions included the following: to enable local health care systems and different group populations to be familiarized with and use mature technological solutions [40,41], to address emerging ethical questions [40], the need to familiarize with tripartite consultations, and to adapt primary care physicians' routines to *ad hoc* meetings [37]. Older rehabilitation patients valued face-to-face contact with their therapist, even when they are very positive about their telerehabilitation experience. This perception of telerehabilitation as complementary rather than a substitute for in-person care indicates that an ideal telerehabilitation service would continue to provide traditional therapy options by interspersing face-to-face contact with distance therapy wherever possible [42].

eHealth was enacted as highly heterogeneous and interdependent concerning composition, roles, and outcome in this cluster. In addition, dynamic adaptation is a keyword that captures services. Services were composed as bespoke, tailored, and interdependent of a variety of human and organizational components.

Videoconferencing as a Means for Administrative Efficiency Improvement

What Works for Whom and in Which Clinical Area?

Video telehealth on outpatient clinic triage nurse workflow consisted of the experimental use of personal computer-based videoconferencing software between home and an outpatient clinic triage office to communicate health questions and concerns that would otherwise be communicated by telephone [43]. For this service, technologies were considered unpredictable, as disrupting existing and well-working routines and dependent on the increased human effort.

Considerations and Proposals for Improvement of Services

Authors referred to a proposal from the involved to extend the scope of involved professionals for assessing a broader spectrum of components to conclude about the effectiveness. In that respect, the service turned out to be more complex than had been assumed before the investigation. This service was disruptive of existing services that were considered as well functioning.

Summary

In chronic neurology, dermatology, and heart disease, thrombolysis for stroke, and palliative home care, videoconferencing was addressed as a singular technological intervention within existing, but varying, organizational arrangements where expert roles were clearly defined. The characteristics were evidence-based advice between specialists and patients, specialists and primary care doctors, and a specialist center and rural hospitals. Success on outcome measures of the effectiveness, convenience, and safety for patients and clinical quality was described. The authors asked for changes in target groups and additional human professional resources for improvement or stabilization. In addition, the authors suggested the appropriate timing and frequency of palliative care teleconsultations, which had shown both positive and negative effects.

In telegenetic services and rehabilitation after total knee replacement, mixed services and novel organizational arrangements were described. Cost savings and the viability of services were demonstrated, and the authors asked for better specification of appropriate outcome and improved knowledge of long-term clinical effects.

In neonatal care, neuromuscular diseases, dermatology and heart disease, telepathology, palliative care, chronic diseases of hypertension, COPD, and cardiovascular diseases, novel organizational arrangements and composite services were described; they were diverse, outreaching, and in need to adapt to contexts and specified challenges. All services were described as conditionally successful but were considered both positive and negative in palliative care services, raising new ethical and relational questions to be addressed for future improvements. It became apparent that videoconferencing posed certain incremental demands for new resources in terms of improved technological features, changes in organizational arrangements, and new professional and ethical efforts. Elderly rehabilitation patients value face-to-face contact with their therapist, even when they are very positive about their telerehabilitation experience.

In triage, videoconferencing was considered as unpredictable and as disrupting existing and well-working routines. Success on the effectiveness was dependent on the increased human effort and extension of involved professionals.

Discussion

Generative Learning and Conceptual Innovations: Adding the 4 D's to the 10 E's of eHealth

We now present the reciprocal translation of the clusters to reach a synthesis by suggesting and arguing for concepts. To further discuss the concepts, we first connect some of the findings to the 10 E's of eHealth and 2 normative statements on what eHealth should be.

The 10 E's of eHealth and Some Characteristics From the Results

Below, Eysenbach's 10 E's of eHealth are listed, along with some characteristics from the results:

1. *Efficiency*: all services in clusters 1 and 4 responded to this characteristic either as goals or as achievements. In clusters 2 and 3, the efficiency was less emphasized.
2. *Enhancing the quality of care*: all services responded to this characteristic as goals and partly as achievements, but the quality was considered very heterogeneously and was ambiguous for palliative care.
3. *Evidence-based*: videoconferencing was partly established on the basis of evidence but also as emerging services, taking one step at a time. Evidence production is a returning challenge in the field as services are adaptive and not always controlled interventions. This is an important issue, which is not addressed in this paper. Perhaps, “early procurement with evidence generation” using participatory approaches

as proposed in Health Technology Assessment (HTA) literature is a way to proceed [44].

4. *Empowerment of consumers and patients*: this was a dedicated goal partly for cluster 2 and for all services in cluster 3.
5. *Encouragement of a new relationship between patients and health professionals toward a true partnership*: this was explicitly stated as a dedicated goal for family caregivers and partly for patients in palliative care.
6. *Education of physicians through online sources*: as described for the telepathology network.
7. *Enabling information exchange and communication in a standardized way between health care establishments*: all services included this as a goal and partly as a characteristic.
8. *Extending the scope of health care beyond its conventional boundaries*: as described for the home care services in palliative care.
9. *Ethics*: eHealth involves new forms of the patient–physician interaction and poses new challenges and threats to ethical issues such as online professional practice, informed consent, privacy, and equity issues; this was addressed especially in palliative care.
10. *Equity*: to make health care more equitable is one promise of eHealth, but at the same time, there is a considerable threat that eHealth may deepen the gap between the “haves” and “have-nots.” Equity was addressed in chronic neurological care, dermatology, and heart disease for thrombolysis, neonatal care, and telepathology.

In addition to these 10 essential E’s, Eysenbach [18] also proposed that eHealth should be:

1. *Easy-to-use*: Our results indicate that this is still a goal and there is still work to bedone, as improvements are requested. Easier use was specially requested for the triage workflow.
2. *Entertaining (no one will use something that is boring!) and exciting*: this subject was not addressed in any of the services, but videoconferencing was considered to interrupt existing and well-working routines in triage.

Adding the 4 D’s to the Concept of eHealth

The quests for different new resources and organizational arrangements revealed videoconferencing as dependent, demanding, and in need to customize to reach goals.

The following 4 D’s can be interpreted as generalizable concepts from our review and serve as input to generative learning: (inter) Dependent; Differentiated across services and temporal lines; Dynamic; and Demanding. The 4 D’s are not easily distinguished as they are also interdependent. We organized the discussion by first coupling the 2 D’s (inter)-dependency and differentiated, and then demanding and dynamic. The former 2 refer to the existing characteristics and the latter to assessments of the existing and future prospects. We interpret these 4 D’s as enacted both within and across clusters and have combined clusters 1 and 4 and 2 and 3 for substantiation. In addition, we propose that eHealth should be:

1. Ethical in that users’ interests should be respected
2. Not harmful by increasing the symptom burden

(Inter) Dependent and Differentiated Services in Clusters 1 and 4

All papers in clusters 1 and 4 [28-32] and [43] were addressed as relatively singular and controlled interventions addressing effective management or coordination. In addition, standardized communication and decision support were expected to improve the clinical quality and patient satisfaction. Considerations about what was needed to sustain services pointed to additional human and organizational resources and more specified solutions. These considerations point to dependent services.

Differences were displayed in the characteristics of the conditions that were considered, as 2 papers addressed singular clinical conditions [28,30], 1 multiple conditions [29], and 1 triage and workflow [43]. In addition, 3 of the 4 papers described services within fixed but different organizational structures—a specialist (center) to local or primary care professionals or patients and with fixed location of the equipment [28-30]. One paper described a network sharing a “virtual consultations room” and differences in results of the intervention for different local sites. Thus, stakeholders indicated dependency and differentiation determined by local variables to explain results [30].

These papers, at the outset considering videoconferencing as a relatively singular intervention, by implication concluded that each service was composite, as well as differentiated across services. Likewise, all services were interdependent, as the inclusion of specific organizational and human resources were considered crucial for continued success. Diversity across services and temporal lines, as well as interdependencies among technological, socioeconomic, and human components, were enacted.

Dynamic and Demanding Services in Clusters 1 and 4

One paper reported negative effects in terms of reduced efficiency. The authors suggested that to conclude, additional organizational aspects like physician time use should be considered. The services played out as dynamic and emerging. For the service to succeed, additional professionals were demanded, as well as further investigations of the impact on the overall clinic workflow [43]. Three papers concluded that the service provided was valuable, but reported the need for additional professional nurses or a health technician to improve the services [28-30]. One reported the need to address issues of increased symptom burden in palliative care [32]. Their assessment pointed both to the interdependency of human resources and demanding and dynamic services. All authors suggested that more knowledge of human (diverse patient populations and competence) and new organizational arrangements was necessary for improvement and more insight.

Videoconferencing was enacted as demanding and dynamic in terms of pressure to incrementally increase human and organizational resources for continued improvement, as well as addressing issues of the increased symptom burden.

(Inter) Dependent and Differentiated Services in Clusters 2 and 3

Common features deriving from this cluster aligned with the former discussions. A closer look at the papers substantiates composite, diverse, and interdependent services by the inclusion of new outreach and additional technological components. Five papers described services provided to patients' homes [36,38-41]. One of these reported advantages of introducing tablet computers to palliative care patients at home [40]. Two papers included monitoring technologies in homes [36,39].

Regarding outcome, the majority reported multiple outcome that substantiates differences across services. Two papers pointed to information exchange, standardized communication, reduced hospital admissions, reduced home visits, and clinical impact [36,37], and one described transcending of institutional walls, clinical quality, and technologized but intimate patient and palliative care specialists' relationships. The authors further considered the opportunity for patients to co-design their own care within the comfort of their homes [40]. While one addressed economic issues (cost-effectiveness or reduction) and quality of life [39], another mentioned retention and recruitment of surgeons to remote hospitals and reduced isolation [37]. In addition, one paper addressed changes in delivery patterns with the intention to refer high-risk newborns from local to regional perinatal centers [35]. One paper addressed new relations between participants and shared decision making by stimulating the integration of primary and specialist palliative care [41].

Technologies were interdependent with peripheral technological equipment involving tablet and monitoring devices and with relational and organizational components as expressed through outreach to patients' homes, shared decision making, and intimate relationships.

Dynamic and Demanding Services in Clusters 2 and 3

The key to further success of telepathology was to maintain and develop a strong communication plan and highly coordinated effort among surgeons, pathologists, stakeholders, laboratory staff and biomedical, administrative, and information technology support teams working in different sites [37]; this key indicated both interdependencies, dynamic development and a demand for additional efforts to obtain objectives.

In palliative care, new ethical questions appeared, which needed attention to meet goals [38,40,41]. The need for additional education, consultation, and guideline dissemination was claimed in neonatal intensive care [35]. In these clusters, the normative aspects also played out—eHealth should be ethical and not serve to increase symptom burden.

Summary

The analysis showed that services were highly interdependent, composite, and diverse across services and temporal lines, as compositions of videoconferencing and peripheral equipment, organizational and human resources, purposes and goals differed between them. They were also demanding in terms of pressure to incrementally increase or change human and organizational resources for continued improvement. The services were dynamic in that new developments were anticipated for continuous development, ranging from dedicated organizational components to strong communication plans involving all stakeholders. Clearly, the 4 D's were highly recognizable and striking in the clusters.

Limitations

To provide a complete account of videoconferencing services in clinical practices in the period covered, the review has limitations in the number of databases that were searched, the search terms, and the 7-year temporal limitation; this was, however, a deliberate decision because we wanted a quick view of recent service configurations. It can be expected, however, that broader search criteria, more databases, and a longer time span would provide even more differentiation and heterogeneity. In that sense, this review provides an informative content that responds to the ideas of realist review. Furthermore, the review was limited in that we did not provide a detailed account of working mechanisms, the how, of different practices.

Comparison With Prior Work and Pragmatic Implications

A systematic review from 2015 addressed technical characteristics of video consultations and ways these had changed over time because of the rapid advancement of information and communication technology [45]. The most widely used hardware described for videoconferencing was dedicated videoconference codecs and personal computers (desktop, laptop, or notebook). In addition, the review reported that the usage of mobile or smartphones for clinical videoconferencing started in 2005 and concluded that this could be an early indication of a marked change from fixed, dedicated hardware (eg, codecs or personal computers) toward ubiquitous devices (eg, smartphones or tablets). The shift to tablets in palliative care supports this change.

In a systematic review of telepathology services from 2016, Farahani et al concluded that mechanisms of success for international telepathology services were efficient workflow, dedicated information technology staff, continuous maintenance, financial incentives, ensuring that all stakeholders were satisfied, and value-added clinical benefit to patient care [46]. Our review has contributed to in-depth knowledge of aspects of workflow, dedicated staff, and value-added clinical benefit in specified clinical settings. What counted as an efficient workflow, dedicated staff, ensuring satisfaction and value-added clinical benefit was highly heterogeneous.

Conclusions

Videoconferencing in clinical practices was enacted as heterogeneous and with bespoke prerequisites and implications. Services were dynamic, differentiated concerning the content and considerations of quality, and adaptive along temporal lines. They were made to work from ongoing demand for fresh resources, making them interdependent. The 4 D's—Dynamic, Differentiated, Demanding, and (inter) Dependent—serve as a pragmatic add-on to Eysenbach's 10 E's of eHealth. Questions concerning the outcome of specified balances between standardization and customization in clinical settings should be addressed in future research, as well as the emerging dual character of outcome—services being considered both “good” and “bad.”

Acknowledgments

This paper is a result of 2 projects publicly funded by the Regional Health Authorities of North Norway in collaboration with the Norwegian Centre for eHealth Research: “Modelling Videoconference Collaboration” [47] and “HTA for ICT” [48]. We are grateful to Eivind Bergman for help with the Excel overview of results and to research librarian Karianne Lind for the search and editing process of references.

Authors' Contributions

AGE designed the study and the electronic form. AGE, AHH, and TSB defined the selection criteria, read papers, and extracted data using the electronic form. AGE wrote the first version of the paper and edited later versions. All authors took part in the analysis, edited the paper for scientific content, and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Electronic proforma.

[MHT File, 217KB - jmir_v20i10e282_app1.mht]

Multimedia Appendix 2

Included primary studies.

[XLSX File (Microsoft Excel File), 63KB - jmir_v20i10e282_app2.xlsx]

References

1. TechTarget. SearchUnified Communications. 2016 Jun 01. What is videoconference (videoconferencing) URL: https://searchunifiedcommunications.techtarget.com/definition/video-conference [accessed 2018-09-27] [WebCite Cache ID 72k91JGRc]
2. Wittson CL, Affleck DC, Johnson V. Two-way television in group therapy. Ment Hosp 1961;12:22-23. [Medline: 14007789]
3. Whitten P, Holtz B, Laplante C. Telemedicine: What have we learned? Appl Clin Inform 2010;1(2):132-141 [FREE Full text] [doi: 10.4338/ACI-2009-12-R-0020] [Medline: 23616832]
4. Hilty D, Yellowlees P, Cobb H, Bourgeois J, Neufeld JD, Nesbitt T. Models of telepsychiatric consultation--liaison service to rural primary care. Psychosomatics 2006;47(2):152-157. [doi: 10.1176/appi.psy.47.2.152] [Medline: 16508028]
5. Heath B, Salerno R, Hopkins A, Hertzig J, Caputo M. Pediatric critical care telemedicine in rural underserved emergency departments. Pediatr Crit Care Med 2009 Sep;10(5):588-591. [doi: 10.1097/PCC.0b013e3181a63eac] [Medline: 19451850]
6. Levine SM, Gorman M. "Telestroke": the application of telemedicine for stroke. Stroke 1999 Feb;30(2):464-469. [Medline: 9933289]
7. Whitten PL, Buis L. Use of telemedicine for haemodialysis: perceptions of patients and health-care providers, and clinical effects. J Telemed Telecare 2008;14(2):75-78. [doi: 10.1258/jtt.2007.070411] [Medline: 18348752]
8. Burdick A. Teledermatology: extending specialty care beyond borders. Arch Dermatol 2007 Dec;143(12):1581-1582. [doi: 10.1001/archderm.143.12.1581] [Medline: 18087012]
9. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): a systematic review of published definitions. J Med Internet Res 2005;7(1):e1 [FREE Full text] [doi: 10.2196/jmir.7.1.e1] [Medline: 15829471]
10. Greenhalgh TJ, 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]
11. May C. A rational model for assessing and evaluating complex interventions in health care. BMC Health Serv Res 2006 Jul 07;6:86 [FREE Full text] [doi: 10.1186/1472-6963-6-86] [Medline: 16827928]
12. Rip A, Kulve H. Constructive Technology Assessment and Socio-Technical Scenarios. In: Fisher E, Selin C, Wetmore JM, editors. Presenting Futures: The Yearbook of Nanotechnology in Society, Vol 1. Dordrecht: Springer; 2008:49-70.
13. Fatehi F, Armfield NR, Dimitrijevic M, Gray LC. Clinical applications of videoconferencing: a scoping review of the literature for the period 2002-2012. J Telemed Telecare 2014 Oct;20(7):377-383. [doi: 10.1177/1357633X14552385] [Medline: 25399998]

14. Granja C, Janssen W, Johansen MA. Factors Determining the Success and Failure of eHealth Interventions: Systematic Review of the Literature. J Med Internet Res 2018 May 01;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
15. Grant M, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Info Libr J 2009 Jun;26(2):91-108 [FREE Full text] [doi: 10.1111/j.1471-1842.2009.00848.x] [Medline: 19490148]
16. Strauss A, Corbin JM. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. Thousand Oaks, California: Sage Publications, Inc; 2015.
17. Walsh D, Downe S. Meta-synthesis method for qualitative research: a literature review. J Adv Nurs 2005 Apr;50(2):204-211. [doi: 10.1111/j.1365-2648.2005.03380.x] [Medline: 15788085]
18. 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]
19. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
20. Pagliari C, Sloan D, Gregor P, Sullivan F, Detmer D, Kahan JP, et al. What is eHealth (4): a scoping exercise to map the field. J Med Internet Res 2005;7(1):e9 [FREE Full text] [doi: 10.2196/jmir.7.1.e9] [Medline: 15829481]
21. Greenhalgh T, Papoutsi C. Studying complexity in health services research: desperately seeking an overdue paradigm shift. BMC Med 2018 Dec 20;16(1):95 [FREE Full text] [doi: 10.1186/s12916-018-1089-4] [Medline: 29921272]
22. Jones R, Rogers R, Roberts J, Callaghan L, Lindsey L, Campbell J, et al. What is eHealth (5): a research agenda for eHealth through stakeholder consultation and policy context review. J Med Internet Res 2005 Nov 10;7(5):e54 [FREE Full text] [doi: 10.2196/jmir.7.5.e54] [Medline: 16403718]
23. Ahern DK, Kreslake JM, Phalen JM. What is eHealth (6): perspectives on the evolution of eHealth research. J Med Internet Res 2006 Mar 31;8(1):e4 [FREE Full text] [doi: 10.2196/jmir.8.1.e4] [Medline: 16585029]
24. Boogerd EA, Arts T, Engelen LJ, van de Belt T. "What Is eHealth": Time for An Update? JMIR Res Protoc 2015 Mar 12;4(1):e29 [FREE Full text] [doi: 10.2196/resprot.4065] [Medline: 25768939]
25. Pawson R, Greenhalgh T, Harvey G, Walshe K. Realist review–a new method of systematic review designed for complex policy interventions. J Health Serv Res Policy 2005 Jul;10 Suppl 1:21-34. [doi: 10.1258/1355819054308530] [Medline: 16053581]
26. Law J. After method: Mess in social science research. London: Routledge, Taylor and Francis Group; 2004.
27. Law J, Urry J. Enacting the social. Economy and Society 2011 Feb 08;33(3):390-410. [doi: 10.1080/0308514042000225716]
28. Davis LE, Coleman J, Harnar J, King MK. Teleneurology: successful delivery of chronic neurologic care to 354 patients living remotely in a rural state. Telemed J E Health 2014 May;20(5):473-477. [doi: 10.1089/tmj.2013.0217] [Medline: 24617919]
29. Johansson A, Lindberg I, Söderberg S. Patients' Experiences with Specialist Care via Video Consultation in Primary Healthcare in Rural Areas. Int J Telemed Appl 2014;2014:1-7 [FREE Full text] [doi: 10.1155/2014/143824] [Medline: 25243009]
30. Agarwal S, Day DJ, Sibson L, Barry PJ, Collas D, Metcalf K, et al. Thrombolysis delivery by a regional telestroke network--experience from the U.K. National Health Service. J Am Heart Assoc 2014 Feb 26;3(1):e000408 [FREE Full text] [doi: 10.1161/JAHA.113.000408] [Medline: 24572251]
31. Akiyama H, Hasegawa Y. A trial case of medical treatment for primary headache using telemedicine. Medicine (Baltimore) 2018 Mar;97(9):e9891 [FREE Full text] [doi: 10.1097/MD.0000000000009891] [Medline: 29489688]
32. Hoek PD, Schers HJ, Bronkhorst EM, Vissers KCP, Hasselaar JGJ. The effect of weekly specialist palliative care teleconsultations in patients with advanced cancer -a randomized clinical trial. BMC Med 2017 Dec 19;15(1):119 [FREE Full text] [doi: 10.1186/s12916-017-0866-9] [Medline: 28625164]
33. Kubendran S, Sivamurthy S, Schaefer G. A novel approach in pediatric telegenetic services: geneticist, pediatrician and genetic counselor team. Genet Med 2017 Dec;19(11):1260-1267 [FREE Full text] [doi: 10.1038/gim.2017.45] [Medline: 28471436]
34. Fusco F, Turchetti G. Telerehabilitation after total knee replacement in Italy: cost-effectiveness and cost-utility analysis of a mixed telerehabilitation-standard rehabilitation programme compared with usual care. BMJ Open 2016 Dec 17;6(5):e009964 [FREE Full text] [doi: 10.1136/bmjopen-2015-009964] [Medline: 27188803]
35. Hall R, Hall-Barrow J, Garcia-Rill E. Neonatal regionalization through telemedicine using a community-based research and education core facility. Ethn Dis 2010;20(1 Suppl 1):136-140 [FREE Full text] [Medline: 20521402]
36. Zamarrón C, Morete E, González F. Telemedicine system for the care of patients with neuromuscular disease and chronic respiratory failure. Arch Med Sci 2014 Oct 27;10(5):1047-1051 [FREE Full text] [doi: 10.5114/aoms.2014.46223] [Medline: 25395959]
37. Têtu B, Perron E, Louahlia S, Paré G, Trudel M, Meyer J. The Eastern Québec Telepathology Network: a three-year experience of clinical diagnostic services. Diagn Pathol 2014;9 Suppl 1:1-5 [FREE Full text] [doi: 10.1186/1746-1596-9-S1-S1] [Medline: 25564940]

38. Oliver DP, Albright DL, Kruse RL, Wittenberg-Lyles E, Washington K, Demiris G. Caregiver evaluation of the ACTIVE intervention: "it was like we were sitting at the table with everyone". Am J Hosp Palliat Care 2014 Jun;31(4):444-453 [FREE Full text] [doi: 10.1177/1049909113490823] [Medline: 23713130]
39. Papadopoulos H. Tile-ippokratis: the experience of an ehealth platform for the provision of health care services in the island of chios and cyprus. Int J Telemed Appl 2010;2010 [FREE Full text] [doi: 10.1155/2010/357156] [Medline: 20871664]
40. van Gurp J, van Selm M, Vissers K, van Leeuwen E, Hasselaar J. How outpatient palliative care teleconsultation facilitates empathic patient-professional relationships: a qualitative study. PLoS One 2015;10(4):e0124387 [FREE Full text] [doi: 10.1371/journal.pone.0124387] [Medline: 25902263]
41. van Gurp J, van Selm M, van Leeuwen E, Vissers K, Hasselaar J. Teleconsultation for integrated palliative care at home: A qualitative study. Palliat Med 2016 Mar;30(3):257-269. [doi: 10.1177/0269216315598068] [Medline: 26269323]
42. Shulver W, Killington M, Morris C, Crotty M. 'Well, if the kids can do it, I can do it': older rehabilitation patients' experiences of telerehabilitation. Health Expect 2017 Feb;20(1):120-129 [FREE Full text] [doi: 10.1111/hex.12443] [Medline: 26890025]
43. Cady RG, Finkelstein SM. Mixed-methods approach for measuring the impact of video telehealth on outpatient clinic triage nurse workflow. Comput Inform Nurs 2013 Sep;31(9):439-449 [FREE Full text] [doi: 10.1097/01.NCN.0000432126.99644.6c] [Medline: 24080753]
44. Facey K, Henshall C, Sampietro-Colom L, Thomas S. Improving the effectiveness and efficiency of evidence production for health technology assessment. Int J Technol Assess Health Care 2015 Jan;31(4):201-206 [FREE Full text] [doi: 10.1017/S0266462315000355] [Medline: 26646858]
45. Fatehi F, Armfield NR, Dimitrijevic M, Gray LC. Technical aspects of clinical videoconferencing: a large scale review of the literature. J Telemed Telecare 2015 Apr;21(3):160-166. [doi: 10.1177/1357633X15571999] [Medline: 25697489]
46. Farahani N, Riben M, Evans AJ, Pantanowitz L. International Telepathology: Promises and Pitfalls. Pathobiology 2016;83(2-3):121-126 [FREE Full text] [doi: 10.1159/000442390] [Medline: 27101287]
47. Bergmo T, Ekeland AG. Modelling VC cooperation: conditions, mechanisms and outcomes – a multi methodological study. Norway: Norwegian Centre for E-health Research; 2013. Modelling Video Conference Collaboration URL: https://tinyurl.com/yb77te7g [accessed 2017-07-13] [WebCite Cache ID 6rvErlPj0]
48. Ekeland AG. Health Technology Assessment (HTA) for ICT. Norway: Norwegian Centre for E-health Research; 2016. URL: https://ehealthresearch.no/en/projects/health-technology-assessment [accessed 2017-07-13] [WebCite Cache ID 6rvFZxG8z]

Abbreviations

COPD: chronic obstructive pulmonary disease
eHealth: electronic health
HTA: health technology assessment

Edited by G Eysenbach; submitted 19.07.17; peer-reviewed by J Shaw, T Jamieson, I Montagni, F Fatehi; comments to author 23.11.17; revised version received 22.03.18; accepted 28.07.18; published 25.10.18.

Please cite as:
Ekeland AG, Hansen AH, Bergmo TS
Clinical Videoconferencing as eHealth: A Critical-Realist Review and Qualitative Meta-Synthesis
J Med Internet Res 2018;20(10):e282
URL: http://www.jmir.org/2018/10/e282/
doi:10.2196/jmir.8497
PMID:30361197

Viewpoint

Web-Based Self-Management Programs for Bipolar Disorder: Insights From the Online, Recovery-Oriented Bipolar Individualised Tool Project

Kathryn Fletcher[1], PhD, MClinPsychol, BSc (Hons); Fiona Foley[1], BSc (Hons), BHSc; Greg Murray[1], PhD, MClinPsychol, BA (Hons), BSc (Hons), Grad Dip (Gestalt Therapy)

[1] Centre for Mental Health, Faculty of Health, Arts and Design, Swinburne University of Technology, Melbourne, Australia

Corresponding Author:
Kathryn Fletcher, PhD, MClinPsychol, BSc (Hons)
Centre for Mental Health
Faculty of Health, Arts and Design
Swinburne University of Technology
PO Box 218, Hawthorn, VIC
Melbourne, 3122
Australia
Phone: 61 3 9214 8900
Email: kfletcher@swin.edu.au

Abstract

Background: Bipolar disorder (BD) is a complex, relapsing mood disorder characterized by considerable morbidity and mortality. Web-based self-management interventions provide marked opportunities for several chronic mental health conditions. However, Web-based self-management programs targeting BD are underrepresented compared with programs targeting other psychiatric conditions.

Objective: This paper aims at facilitating future research in the area of self-management of BD and draws insights from the development of one such intervention—the Online, Recovery-Oriented Bipolar Individualised Tool (ORBIT)—that is aimed at improving the quality of life of people with BD.

Methods: We have discussed the opportunities and challenges in developing an engaging, evidence-based, safe intervention within the context of the following three nested domains: (1) intervention development; (2) scientific testing of the intervention; and (3) ethical framework including risk management.

Results: We gained the following insights across the three abovementioned overlapping domains: Web-based interventions can be optimized through (1) codesign with consumers with lived experience to ensure relevance and appropriateness to the target audience; (2) novel content development processes that iteratively combine evidence-based information with lived experience perspectives, capitalizing on multimedia (eg, videos) that the digital health space provides; and (3) incorporating Web-based communities to connect end users and promote constructive engagement by access to a Web-based coach.

Conclusions: Self-management is effective in BD, even for those on the more severe end of the spectrum. While there are challenges to be aware of, guided self-management programs, such as those offered by the ORBIT project, which are specifically developed for Web-based delivery provide highly accessible, engaging, and potentially provocative treatments for chronically ill populations who may otherwise have never engaged with treatment. Key questions about engagement, effectiveness, and cost-effectiveness will be answered by the ORBIT project over the next 18 months.

(J Med Internet Res 2018;20(10):e11160) doi:10.2196/11160

KEYWORDS

Web-based intervention; bipolar disorder; self-management

Introduction

Background

The well-documented strengths of Web-based delivery for psychological interventions—flexible access across space and time, low cost, and potential for personalization [1-4]—have been bolstered with growing evidence that clinical effect sizes for many common clinical presentations are as large as those achieved by traditional face-to-face interventions [5-7]. As more and more of human life is mediated through technology, it is, perhaps, not surprising that the archetypal personal encounter of psychotherapy is also finding its feet on the Web.

The aim of this paper is to accelerate progress toward the next generation of Web-based interventions by critically reviewing the experience of one, very particular, Web-based treatment development process. The overarching aim of the Online, Recovery-Oriented Bipolar Individualised Tool (ORBIT) project was to develop and test a Web-based self-management intervention for people with bipolar disorder (BD) to improve their quality of life [8,9].

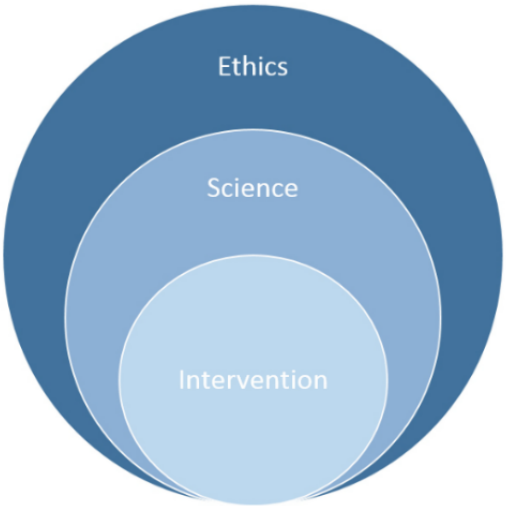
BD is a complex, relapsing mood disorder characterized by considerable morbidity and mortality. Functioning levels vary widely between and within individuals with BD, presenting a challenge for services organized primarily around management of chronically low-functioning patients [10,11]. Web-based self-management programs for this group represent a unique opportunity to address this need [12]. Unfortunately, Web-based self-management programs targeting BD are underrepresented relative to programs targeting other psychiatric conditions [13]. Developing an engaging, evidence-based, safe Web-based self-management program for individuals with BD presents not only special challenges but also opportunities.

We hope that the information presented here will be of use to others developing and testing Web-based psychological interventions. While we focus on a Web-based intervention for BD as a guiding example, our learnings are generalizable to interventions for other mental health conditions. Insights are consequently organized in the following three domains: (1) development of the intervention; (2) development and conduct of a rigorous scientific test of the intervention's efficacy and mechanisms of action; and (3) development of a best-practice risk management and ethical framework for the trial (and ultimately for the roll-out of the intervention). From an overarching project perspective, it is useful to think about these three domains as nested (Figure 1).

The Online, Recovery-Oriented Bipolar Individualised Tool Project

Our international team is currently in the latter phases of recruitment for the randomized controlled trial (RCT) component of the ORBIT project. The trial (registered 23 June, 2017, ClinicalTrials.gov: NCT03197974) is funded by the Australian National Health and Medical Research Council and was reviewed and approved by Swinburne University of Technology Human Research Ethics Committee (2016/289). As detailed in the protocol paper [9], recruitment is primarily via open social media sites (eg, International Bipolar Foundation Facebook site). Main inclusion criteria are as follows: a diagnosis of BD (confirmed via a phone-administered structured diagnostic interview), no current mood episode, history of ≥10 mood episodes, no current psychotic features or active suicidality, and under the care of a medical practitioner. The trial compares two contrasting interventions, both referred to as self-management programs (Mindfulness vs Psychoeducation) for BD [9]. The Web-based programs aim to improve the quality of life in those with "late stage" BD (defined as ≥10 mood episodes). The programs (accessible via personal computer, tablet, and mobile phone) are brief (4 modules delivered over 5 weeks), self-paced and tailored to BD. They incorporate a range of multimedia components to maximize engagement and motivation: videos of consumers with the lived experience and clinicians (shaping program content), audio files for practicing learned concepts, interactive exercises, quizzes, static images, and downloadable PDF content for further learning opportunities. Guided support is offered via once-weekly asynchronous messages (from trained coaches); peer support is offered via moderated forums and the ability for users to connect privately with each other via a secure-messaging system embedded in the program.

Figure 1. Project domains.



The programs are intended to be highly interactive; users are encouraged to track and monitor well-being via an embedded tracking tool, complete interactive exercises and reflect on their participation as they complete each module, contribute to forums, connect with other users (fostering social support), and message their coach for assistance. Furthermore, engagement and adherence are encouraged via coach messages, seeded forum posts, and cognitive behavioral principles to facilitate the practice of skills in everyday life.

Unlike the standard approach of adapting validated face-to-face psychological interventions for Web-based delivery, the programs were bespoke Web-based interventions. Content was driven by evidence-based psychological treatments, offered via the following two arms: (1) Psychoeducation for BD (serving as the active control condition) and (2) Mindfulness for BD (incorporating elements of mindfulness-based cognitive behavioral therapy, self-compassion, and Acceptance and Commitment Therapy). Full details of the ORBIT project, including the rationale for targeting “late stage” BD are outlined in the protocol paper [9].

Insights from the ORBIT project are now considered in terms of strategies to optimize the intervention, scientific considerations for testing the intervention, and the framework that was developed to optimize ethical conduct of the research and delivery of the intervention.

Key Learnings: Opportunities and Challenges

Intervention

Here the intervention offered in the ORBIT project refers to the Web-based self-management programs. A number of strategies were used to optimize these programs, now detailed.

Consumer Input

Consumers with lived experience of BD were involved in all phases of program development, aligning with the consumer-based participatory research approach [14]. Consumer feedback from the pilot phase [8] guided content development for the current iteration of ORBIT. Intervention content is largely driven by the consumer voice, captured via videos of those with lived experience of BD. A local consumer advisory group (CAG) was established for the ORBIT project, comprising 10 individuals (6 females and 4 males) diagnosed with BD. Three CAG meetings took place during the development phase. Members played an integral role in providing feedback on content and the website itself (look-and-feel, ease of use, any perceived benefits or barriers) to ensure appropriateness and relevance to those with BD. We additionally consulted our pre-established CAG (located in Canada), part of the Collaborative Research Team to study psychosocial issues in BD (CREST.BD), to ensure that intervention content was appropriate for a broader international audience.

As part of the RCT, qualitative feedback from participants completing the program was collected to (1) guide future developments and (2) provide insights into their level of engagement with the Web-based intervention. These practices extend beyond the usual application of face-to-face intervention for the Web-based format by allowing a largely consumer-driven process. This approach ensures the program is tailored to the population it seeks to serve and aligns with the recovery model focus of empowering consumers via their active involvement in all stages of intervention development and testing.

Use of Multimedia to Develop Content and Explain Concepts

Our international team of researchers, clinicians, and consumers initially developed and piloted program content drawn from mindfulness-based therapies [8]. The program was then extensively revised and extended, iteratively developed via a dynamic interplay of theoretically derived therapeutic content and footage from videos of consumers with lived experience of BD. Content for Web-based interventions is often adapted from a face-to-face psychological intervention for Web-based delivery, as part of a one-step process where the content is largely defined upfront. By contrast, ORBIT content was progressively developed over a 6-month period. The process commenced with the “top-down” development of a semistructured interview schedule, drawing from broad topic areas aligned with mindfulness-based therapeutic approaches and BD psychoeducation. Clinical psychologists and mindfulness practitioners were consulted on topic areas to ensure that questions were grounded in a psychological framework. Next, 12 consumers were recruited to participate in the filming process: consumers were selected on the basis of gender; age; cultural background; and a range of experiences with mindfulness, acceptance-based approaches and self-management strategies (eg, recognizing early warning signs and triggers), ensuring a representative sample that ORBIT participants could relate to. The video process adopted a documentary-style interview, lasting up to 2 hours per person, undertaken with a professional film crew. Consumers were encouraged to speak from personal experience to ensure footage captured the central “consumer voice.” This “bottom-up” generation of program content was balanced within the “top-down” psychological framework—footage was carefully reviewed, and the project team iteratively revisited the planned content and structure of the intervention to maintain this balance. Hours of footage were edited into “snippets” from a range of consumers and combined into short videos (3 minutes on average in length), describing particular skills and experiences. This new type of delivery (documentary-style videos as an engagement strategy for Web-based interventions) ultimately led to a novel way of developing a Web-based intervention and, ultimately, a new intervention. The final stage of development involved fleshing out content from videos, including key summary messages, supplementary text, and MP3 audio files (allowing for in-the-moment experiential practice of newly introduced concepts) to promote learning and encourage skill development. Program content was then reviewed by our local and international CAGs and revised on the basis of their feedback. Overall, feedback indicated that videos successfully captured the consumer voice, providing a “real-world” feel that people could relate to, adding credibility to the program (balancing realistic messages with hope), reducing stigma, and allowing subtle processes to be described in an engaging and personalized way:

> *I like the real experiences shared by those with Bipolar better than those put forward by clinicians. A shared personal lived experience of Bipolar resonates deeply. Whereas I always feel a clinician is regurgitating something from a textbook or re-telling someone else's personal story: they haven't lived it, so they will never truly know.* [Female, age 51 years, bipolar II disorder]

CAG members were consulted on the duration of videos; in general, there was agreement that "shorter" (3-4 minutes) videos were preferable in terms of maintaining interest and engagement, without being too cognitively demanding. This is in accordance with our prior experience in developing video-based content for Web-based interventions for individuals with persisting psychosis as a strategy to promote recovery [15]. Finally, the use of informational PDFs, clinician videos, and instructional videos (eg, consumer or expert walk-through, how to use a mood tracking tool) allows concepts to be easily understood within a user-friendly context. The Web-based environment lends itself to participants revisiting content as needed, at their own pace, to consolidate their understanding and facilitate repeated practice.

Content is guided by psychological principles; thus, consideration of how this can be made engaging, interactive, flexible, and appealing is essential to the success of any Web-based self-management program. Given the cognitive impairments common in those with BD, practical considerations (eg, being able to start, pause, and recommence topic areas; videos of short duration accompanied by transcripts; length of exercises; language; and neutral colors and icons) were carefully considered to ensure ongoing engagement.

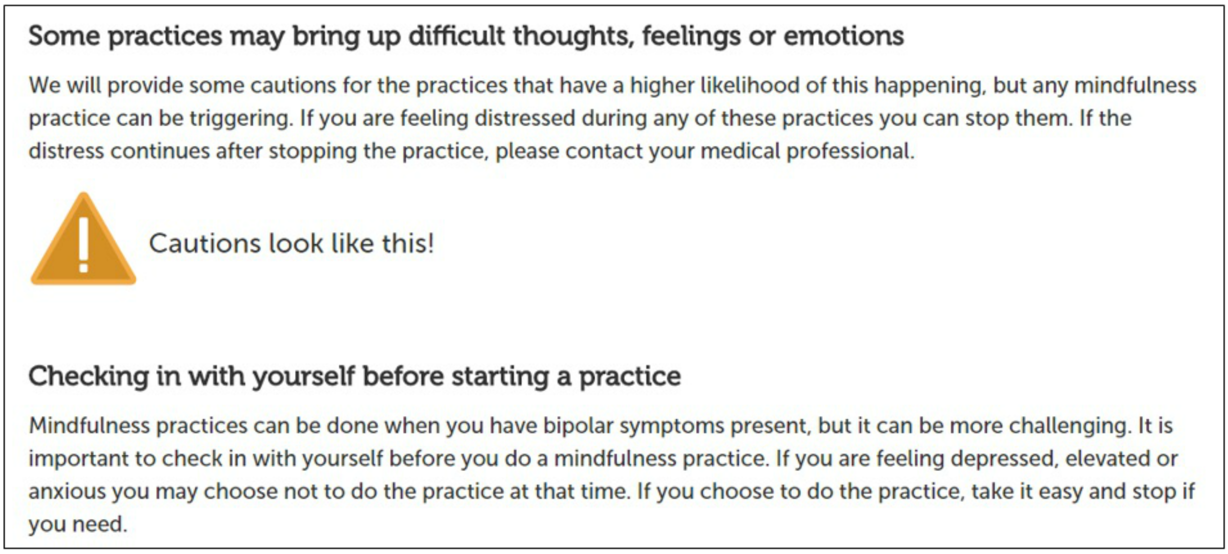
Access to a Web-Based Community

The programs were designed to maximize constructive engagement by participants having access to a Web-based community (other participants, moderated forums), allowing them to link in with social support that may not otherwise have been realized [16]. This can provide a sense of normalization and a way in which they can also support others. The forum facilitates sharing of key learnings as participants navigate through the program, and a secure-messaging system allows participants to foster connections with other users if they so wish (extending social support networks outside of ORBIT). Participants are assigned a Web-based coach as part of the ORBIT community, with asynchronous message support. As overviewed by others [17], support in Web-based interventions for BD and other psychological conditions reduces dropout rates, making them comparable to rates observed in face-to-face therapies. Indeed, motivation to persist with Web-based interventions has been found when needs for relatedness (eg, identifying with other end users and content, support from Web-based coach, forum participation) are satisfied [18].

Clinical Cautions

Our group has been interested in developing novel psychotherapies for BD, drawing on third-wave principles of mindfulness and self-compassion [19]. There have been anecdotal concerns in the literature about potential iatrogenic effects of mindfulness for people with BD [8]. One of the arms in the ORBIT project contains such elements, requiring meticulous attention to ethical issues and risk management. Current mood state is an important factor to consider when introducing those with BD to experiential techniques within a Web-based intervention, given the risk of triggering mood dysregulation. Clinical caution messages (both within the content and incorporated into audio exercises) are used to empower participants to consider whether practicing a particular technique at that moment would be beneficial or should be delayed. For instance, anecdotal reports of body scan exercises that are lengthy (eg, 20 minutes or longer) indicate they can be triggering (of mood episodes) for some; thus, cautions based on current mood state can be particularly useful (Figure 2).

Figure 2. Clinical cautions. Source: the ORBIT program.



Some practices may bring up difficult thoughts, feelings or emotions

We will provide some cautions for the practices that have a higher likelihood of this happening, but any mindfulness practice can be triggering. If you are feeling distressed during any of these practices you can stop them. If the distress continues after stopping the practice, please contact your medical professional.

Cautions look like this!

Checking in with yourself before starting a practice

Mindfulness practices can be done when you have bipolar symptoms present, but it can be more challenging. It is important to check in with yourself before you do a mindfulness practice. If you are feeling depressed, elevated or anxious you may choose not to do the practice at that time. If you choose to do the practice, take it easy and stop if you need.

This type of encouragement allows participants to self-pace through the program, allowing for increased engagement and a sense of autonomy. Participants are encouraged to reflect on exercises, and any adverse events arising that are directly related to intervention content are logged.

Sustainability

A widely recognized problem in internet delivery is transitioning from research-funded development and trialing of an intervention to long-term sustainable delivery outside a research funding environment. Swinburne University of Technology has been at the forefront of improving the sustainability of Web-based interventions by advocating (successfully) for the national recognition of eTherapy hours as integral to the training of Clinical Psychology Masters students (25 eTherapy hours approved by the Australian Psychological Society College of Clinical Psychologists). In the ORBIT trial, and in future potential iterations of the ORBIT website beyond the research phase, the primary personnel support (coaching) is provided by Clinical Psychology Masters students as part of their first internal placement at the university, a model that can be generalized to any psychological interventions offered in Australia.

Science

Randomization and Allocation

A key facet of minimizing bias in treatment trials is ensuring that participants are randomly allocated to comparison conditions [20]. Web-based RCTs benefit from the use of fully automated computer-generated blocked randomization *and* allocation methods, ensuring all aspects of randomization are fully concealed to research personnel. Such methods are superior to traditional RCT methods (eg, sequentially numbered opaque sealed envelopes, central randomization by telephone to a trials office, etc) as no human involvement is required (an airtight method of concealment, reducing bias) and time efficiency is achieved (randomization and allocation can occur within seconds of each other, allowing participants to commence the intervention almost immediately). The latter is particularly important for studies involving those with frequent mood changes (ie, BD) that can occur within hours or days, impacting study data.

Defining “Dose” in the Web-Based Context

A central challenge posed by the Web-based context is how best to define intervention “dose.” An effective “dose” can constitute the level of usage needed for participants to benefit from the use of the program, the extent that the “dose” varies between participants, and participant characteristics that may influence the “dose” that is needed [21].

Participants differ in Web-based usage patterns, with evidence suggesting that less time will be spent on the program than researchers expect [22]. Program usage statistics automatically recorded onsite (eg, number of pages viewed, number of tasks completed, timestamps, etc) are an essential measure of adherence and engagement, providing an objective proxy for “dose” received. Study investigators need to identify program usage statistics of interest during the development stage to communicate this to the website developer (who will incorporate selected variables into data downloads for later analysis). Usage data can automatically be tracked and operationalized in an algorithm combining time on the Web, activity completion, and active engagements with the intervention [23]. However, adherence (and the related construct of engagement) is difficult to characterize and measure—for instance, the proportion of time spent on a particular page does not necessary represent the time participants were actively engaged on that page (they may have been chatting on Facebook or away from their computer). Furthermore, adherence alone does not capture a participant’s entire experience of an intervention [23]; there is growing interest in understanding other ways in which individuals *engage* with Web-based interventions, and the meaning of *engagement* in this context [24]. Currently, minimal consensus exists on the definition and conceptualization of engagement with Web-based interventions. Some view engagement as synonymous with adherence and the opposite of “intervention attrition” or “treatment dropout” [25,26], while others consider it to move beyond mere attendance, incorporating the extent that an individual *actively* participates in a treatment on offer behaviorally, cognitively, and affectively [27,28].

To develop a more comprehensive understanding of adherence and engagement, the ORBIT project will examine both objective and subjective levels of “active” participation; usage statistics will be captured via the website, and in-depth qualitative interviews will be conducted with a subset of participants asking about their usage (online and “offline” practice of skills) and *perceived level of engagement* with the content. Through this process, we will develop an innovative algorithm to quantify the important variables of “minimum dose” and “attrition” in the Web-based context for inclusion in statistical analyses. As discussed in the protocol paper [9], the concept of attrition appears in two sensitivity analyses of the primary outcome. First, per-protocol analyses will be conducted on those receiving a minimal dose of the intervention, with minimal dose to be defined on the basis of the pending algorithm of self-reported and automatically recorded usage variables. Second, intention-to-treat analyses will be repeated with imputation on baseline and relevant postrandomization variables; these variables could include attrition (again defined in the pending algorithm of usage variables) if it is shown to differ between groups. As also outlined in the protocol paper, we will attempt to follow and assess all participants regardless of the level of usage of the site during the intervention period (with the exception of those explicitly discontinuing or being withdrawn on ethical grounds).

Minimizing Nonadherence and Maximizing Engagement

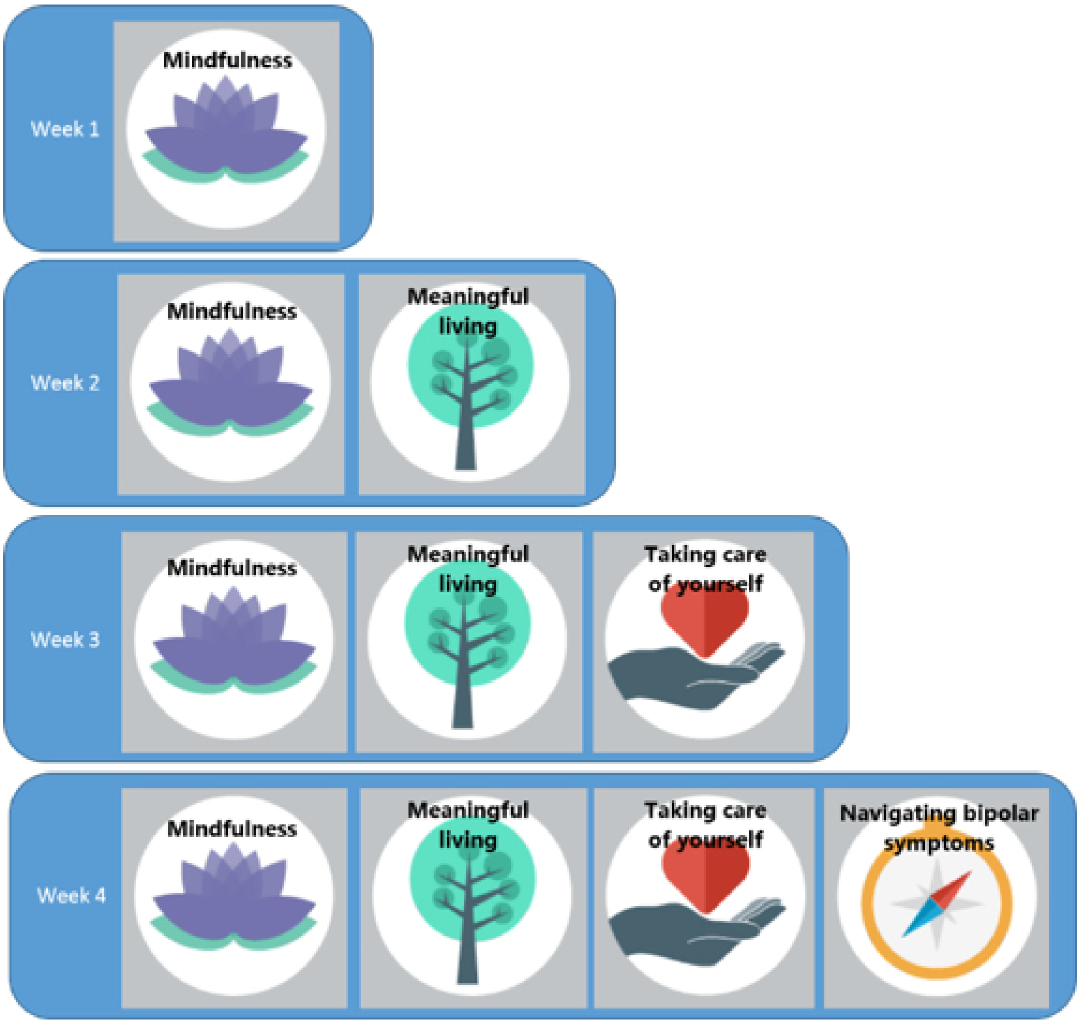
As outlined in our protocol paper [9], the interventions follow best practice in persuasive system design. Three key features known to impact on engagement are utilized: (1) *dialogue support* (praise from coach and forum moderator, email reminders); (2) *social support* (social facilitation through discussion threads in moderated forums); and (3) *primary task support* (best-practice principles for modularization of content, personalization or monitoring of progress, prompted self-monitoring, and rehearsal) [29]. Furthermore, the intervention is brief, and program content is released to each participant sequentially each week in an attempt to pace users as they work their way through the program, while gradually building on knowledge from earlier weeks (with "teaser" messages with respect to upcoming content). We opted against the delivery strategy of a single exposure (receipt of all content in one go), despite some evidence suggesting engagement rates increase when end users have control over how they view content, along with free choice on when they interact with it [30]. Our decision was guided by qualitative consumer feedback from our pilot study [8], indicating some end users felt overwhelmed with the amount of content.

An open approach to navigation within each week's content invites a further challenge—the program is essentially a "set of offerings" rather than "sessions" or "modules" to be completed, catering where possible for different users (eg, those with no experience of mindfulness vs regular meditators; those who have limited time to spend on the program). Unlike manualized treatment programs adapted for the Web-based format, one cannot assume that participants will work their way through each week's content in a sequential manner. This required careful consideration during ORBIT content development, with persuasive technologies (eg, links and "suggestions" embedded within the content) used to prompt participants on how they might best navigate their way through the program, "chunking" of material to allow for shorter sessions times depending on user preferences, and suggestions for skills practice peppered throughout the content in case users did not click through to the "homework" page located at the end of each week's content. Topic areas were crafted in a largely self-contained way, for example, using mindfulness (ranging from beginners exercises to more advanced practices) as a guiding overall skill to integrate topics (Figure 3).

Consumer Involvement

Consumers (referred to as "superusers") with lived experience of BD were employed for the ORBIT project and trained to assist with seeding forum content and facilitating engagement with the content. This strategy brings unique opportunities and challenges. While a rich and lively forum community has developed and participants appreciate and benefit from communication with peers, guiding principles have been developed to manage the ethical and scientific impact of this dynamic environment. The superuser role is defined as nontherapeutic, and a degree of self-disclosure is encouraged.

Figure 3. Online, Recovery-Oriented Bipolar Individualised Tool (ORBIT) topics. Source: the ORBIT program.



From a scientific perspective, superusers are instructed to seed forum content with messages that are specific to the program being offered (eg, mindfulness-related material), avoiding (where possible) cross-contamination with content from the alternate program. Monthly supervision of superusers is undertaken to ensure that these guiding principles are adhered to while balancing the tension of allowing the forums to organically unfold as part of a dynamic Web-based community.

Ethical Framework

Remote delivery considerations associated with Web-based self-management programs include risk-management, participant distress, and legal issues around delivering interventions in different jurisdictions.

Risk Management

Delivering an adjunctive Web-based self-management program for BD (and indeed, any chronic mental health condition with substantial clinical risks) within the context of an international research trial brings a level of risk that requires clear protocols should adverse events arise. Real-time intervention is not always possible or feasible given users are on the Web during different time zones, and its use must be carefully considered.

Motivated by clinical risk-management priorities, as well as principles around generalization of learning through participation, an approach was developed that emphasizes participants' local networks of treatment and care [31]. Specifically, our approach was to explicitly devolve safety and well-being to participants and their local network (treating mental health practitioner and local emergency services). This is achieved in three key ways. First, an inclusion criterion requires participants to have had contact with a mental health professional during the past 12 months, to provide these details to the project team, agree for this professional to be contacted if necessary, and understand that they remain the first point of contact. Mental health professionals are posted a courtesy letter informing them of their clients' participation in the ORBIT project. Second, participants are explicitly made aware (via the consent form, during the sign-up process when speaking to the research team, information on the program websites, coach messages, and forum messages from superusers) that the program is not intended to replace their usual care, does not provide a crisis service, and is not monitored in real-time. Links to emergency resources are, nonetheless, provided on the program websites (eg, *unsuicide.wikispaces.com*). Participants are directed back to their mental health professional as needed throughout their involvement in the ORBIT project. Third, a "red flag" protocol was developed to guide adverse event procedures, based on our experience with other Web-based interventions and websites for BD [32-37] and consultation with the *CREST.BD* CAG. This is detailed in the protocol paper [9]. In essence, the decision tree distinguishes between red flag information (which can arise during research assessments, Web-based questionnaires, forum posts, and messages to coach), suggesting immediate risk of harm for which real-time intervention is feasible or recommended (eg, when active suicidality is identified during a phone assessment), and when it is not (eg, when the research team becomes aware of active suicidality mentioned in a forum post from 48 hours previous). Actions are progressively escalated to senior staff members if required. As an example, forums are moderated by the project manager; inappropriate content is initially flagged (both in terms of risk and potential for triggering other participants) by superusers, prompting appropriate actions. All trial staff are comprehensively trained on these protocol procedures, operating to the guiding principle that participants' local treatment and care networks are not disrupted. As per standard ethical guidelines, participants are withdrawn from the ORBIT project on a case-by-case basis should it be deemed that their well-being is compromised by their participation, and serious adverse events suspected or known to be related to participation in the trial are reported to the local administering ethics committee (Swinburne University of Technology Human Research Ethics Committee).

Participant Distress

Related to risk management, participant distress (whether arising as a direct result of participation in the project or as part of the usual clinical course of the mental health condition) is an ongoing challenge and particularly so for remotely delivered interventions. As for the majority of RCTs (whether Web-based or face-to-face), the ORBIT project includes a structured diagnostic interview, the Mini-International Neuropsychiatric Interview [38] to assess inclusion or exclusion criteria. Structured diagnostic interviews are, by their very nature, highly detailed and require participants to revisit their experiences of distressing symptoms. When conducted face-to-face, interviewers are able to pay attention to nonverbal cues and can manage distress levels sensitively. This poses a particular challenge for interviews conducted over the phone, as for the ORBIT project. Distress associated with participation in the diagnostic interview has been flagged by consumers participating in training for research staff administering the Mini-International Neuropsychiatric Interview, as well as a small proportion of ORBIT project participants. The detailed nature and duration of the baseline interview (lasting up to 2 hours) can be particularly distressing for those who have experienced multiple mood episodes. Solutions to date have included warning participants upfront of the potential for distress arising, interviewers being trained to tune in to participant tone of voice and other verbal cues, offering frequent breaks during the interview and conducting the interview across a few sessions if needed. While structured diagnostic interviews are a standard component of mental health research, sensitivity to participant distress and burden (particularly for those with chronic disorders such as BD) is a key priority. Offering participants the opportunity to debrief following such interviews may be an additional solution for future studies in this space.

Legal Issues

A key issue currently facing the delivery of Web-based "interventions" for mental health concerns the type of intervention that is offered. Interventions claiming to have some therapeutic value (eg, psychological interventions) fall into a gray legal zone, whereby certain jurisdictions require the "therapist" providing the "intervention" to be registered in the state, territory, or country where the client accesses the service from. Indeed, some states could potentially prosecute the remote "therapist" under the state's laws. As clients can access

Web-based programs from any location worldwide, this presents a legal minefield. As a first step to navigate this, the terminology used to describe the intervention must be carefully specified—for example, the ORBIT project does not claim to offer a psychological service, rather a self-management program that complements (but does not replace) usual clinical care. Second, care should be taken in defining what the program is intended to offer—the ORBIT project indicates that improvements in quality of life may result from completing the program. Third, as a guided program where participants have access to a personal coach, coach qualifications and role are made clear upfront. Specifically, the coaching role is nontherapeutic, with the key aim of supporting participants in terms of their *engagement with program content*. The tone and content of coach messages are carefully crafted on the basis of general guiding principles to standardize responses where possible (while remaining "human"), and coaches receive regular supervision. For example, responses tending toward therapeutic statements (eg, advanced empathic insights) were discouraged on two grounds: (1) the asynchronous email communication cannot sustain such a dialogue and (2) these could weaken participants' engagement with their own local therapeutic resources. Finally, as a Web-based program within the context of an RCT, we strategically positioned the ORBIT project as a "single-site" study, governed by a single Human Research Ethics Committee (HREC; in this case, Swinburne HREC), offering services delivered from Australia outlined in a disclaimer present in the website terms and conditions reading, "*The services on the ORBIT website are provided in accordance with Australian laws and health practice standards. You acknowledge and accept that the services may not comply with the laws and standards that apply in the jurisdiction in which you receive the services.* " Many of the legal issues associated with the delivery of Web-based interventions remain unknown at present while university HRECs have not developed internal processes to deal with such issues.

Summary of Learnings

Insights across the three overlapping domains (intervention development, scientific testing, and ethical frameworks) are now summarized. Web-based interventions can be optimized through (1) codesign with consumers with lived experience to ensure relevance and appropriateness to the target audience; (2) novel content development processes that iteratively combine evidence-based information with lived experience perspectives, capitalizing on multimedia (eg, videos) that the digital health space provides; (3) incorporating Web-based communities to connect end users and promote constructive engagement via access to a Web-based coach. The potential iatrogenic effects of particular exercises (eg, those of an experiential nature) within the intervention content must be considered for the target group of interest, with clinical caution messages and self-pacing encouraged to ensure end users move safely (and autonomously) through program content. Finally, sustainability models that are generalizable (eg, personnel support for guided interventions outside of an RCT context) should be considered as part of the development process.

Within the scientific context, while Web-based RCTs offer swift, unbiased randomization and allocation methods that can be fully automated, adherence and engagement with the intervention itself can be difficult to quantify. These concepts require further investigation; study designs should incorporate quantitative and qualitative assessment of adherence and engagement to move beyond automatically captured usage data and develop a richer understanding of how participants interact with Web-based interventions. These learnings can guide persuasive technologies to optimize an individual's experience of the program. As a final method of engaging users, ORBIT seeks to build a dynamic Web-based community driven by consumers with lived experience. The ethical and scientific impact of this environment requires careful consideration; superuser roles require clear definition, seeded forum content should align with the intervention content (avoiding cross-contamination with the control condition), and ongoing supervision is required to allow forums to unfold organically while balancing scientific integrity.

Ethically, clear risk management protocols are required for Web-based self-management programs. Real-time intervention is not always possible, nor feasible, within the context of an international RCT. A guiding principle of the ORBIT project is to explicitly devolve safety and well-being to participants and their existing local care network, respecting clinical risk-management priorities and participant autonomy. While structured diagnostic interviews are central in determining inclusion criteria to RCTs, awareness of participant distress arising from such interviews requires sensitivity, careful training of clinical interviewers, and debriefing where necessary. Finally, careful consideration of legal issues surrounding the delivery of interventions claiming to have "therapeutic value" in different jurisdictions is warranted.

We offer some final insights with respect to the multifaceted skill set requirements for projects such as ORBIT that are broadly applicable to the development of any Web-based intervention. A level of technological expertise is needed to oversee the development, implementation, and evaluation of Web-based self-management programs. Expertise in this space can include knowledge of e-learning, e-communication, health informatics, basic programming skills, and awareness of technological barriers that could deter use (eg, slow internet speed and ability to watch videos), with the overall aim of ensuring that the program is innovative, engaging, feasible, and likely to be effective. As described by others [39], the ability of mental health researchers to enter the world of the Web-based program developer invites valuable opportunities to influence scoping, design, and evaluation. This offers a new skill set to mental health researchers—a new "breed of transdisciplinary experts" [39]—allowing highly innovative and clinically effective electronic mental health programs to be developed.

Discussion

Principal Findings

Moving forward, the digital health space offers multiple opportunities. First, Web-based self-management interventions such as ORBIT could be integrated into a stepped care approach in primary care [40]. Dissemination would occur via Primary Health networks; general practitioners are well placed to identify patients who may benefit from an evidence-based, low-intensity Web-based intervention as a first step. Stepping up or down the treatment pathway would then be determined according to patients' needs and response to treatment. Second, hybrid treatments, where mental health practitioners and patients use Web-based programs in conjunction with face-to-face treatments (eg, within sessions together or as a way of stimulating discussions and promoting continuity of treatment outside of the treatment room), may provide alternative (or complementary) models of care. Practitioners and their patients with serious mental illness have expressed positive views about this model [41], which is currently under evaluation [42]. Third, moderated discussion forums such as those included in ORBIT may serve as stand-alone interventions; these empowered Web-based communities provide a rich, dynamic environment where consumers can exchange mental health information and receive support [43]. The peer discussion boards of the MoodSwings 2.0 Web-based self-help program for BD are currently being evaluated to clarify and maximize the benefits of Web-based discussion [44].

While clear opportunities exist in the digital health space, key challenges remain in terms of delivery and adherence. Legalities surrounding Web-based delivery of mental health interventions across different jurisdictions require urgent attention. Currently, university ethical review boards are either unaware of, or not resourced to address, legal issues arising from geographic jurisdictions other than their own [37]. For review boards to undertake appropriate vetting of clinical trials of Web-based interventions, clear identification of protective factors (eg, participants' privacy and safety, ethical considerations, and risk issues) across jurisdictions is necessary. Transparent models for multinational internet intervention research initiatives are now needed to navigate these legalities. One component of such a model may include clearly informing participants in consenting documents that while their participation has been ethically vetted by only one institution in a given geographic and legal jurisdiction, they remain bound by legal and ethical precedents in their own geographical jurisdiction [37].

Turning to adherence with Web-based interventions, attrition is an ongoing challenge. Unlike face-to-face treatment trials where the common factors (ie, therapeutic relationship) can bolster adherence to the control condition, Web-based treatments (particularly those without guided support) pose different challenges. Research trial investigators must now make the choice to address attrition by making the engagement features of the control condition comparable to those of the preferred condition. The intriguing notion of the "therapeutic relationship" in nonguided Web-based interventions requires further exploration both in terms of attrition and treatment outcomes [45].

This discussion of insights around developing and testing a novel Web-based intervention for BD was organized in terms of three nested foci: the intervention, the science, and the ethical framework (see Figure 1). We propose that this organizing scheme is useful for future efforts in this space, particularly because it helps illuminate tensions between these three critical goals of any eTherapy project. The ORBIT project required decisions about, for example, (1) the *scientific* preference to offer a standard "no frills" control condition versus the *intervention-level* need to have best-practice engagement in the control condition; (2) the *scientific* preference to make findings as generalizable as possible versus the *ethical* need to constrain participation to minimize the risk of adverse events; and (3) the preference to have relatively unconstrained discussion on the forums to optimize impact of the *interventions* versus the *scientific* goal of stimulus control and the *ethical* goal of minimizing triggering statements for other participants. The ultimate development or testing of any Web-based intervention rests heavily on the contextualized, procedural positions taken on these multifaceted issues, reminding us again of the substantial gap between the simple content of any psychological intervention and its instantiation in a Web-based delivery platform.

Limitations

This paper aimed to support future investigation of Web-based interventions for mental health conditions by describing the minutiae of the clinical, scientific, and ethical decision making underpinning one particular Web-based trial. A potential limitation of the paper, then, is the extent to which insights from the ORBIT trial are specific to this population (patients with BD), this intervention (brief novel self-help strategies), or this outcome variable (quality of life). We encourage readers to be attentive to these particulars as they draw generalizations for their own innovative Web-based interventions.

Conclusion

Technology allows for highly interactive and engaging programs that empower participants to manage their mental health. This departure from the care model can prompt clinician insecurity (treatments that work are arguably provocative and, therefore, potentially destabilizing); however, it should not be a barrier to offering the intervention to consumers. As overviewed by others [12,17], there is evidence that self-management is effective in BD, with those on the more severe end of the spectrum still able to learn to self-manage and take control of their lives. While there are challenges to be aware of, guided self-management programs such as those offered by the ORBIT project that are specifically developed for Web-based delivery provide highly accessible, engaging, and potentially provocative treatments for chronically ill populations who may otherwise have never engaged with treatment. Key questions about engagement, effectiveness, and cost-effectiveness will be answered by the ORBIT project over the next 18 months.

Acknowledgments

This project was supported by the Australian National Health and Medical Research Council project grant (APP1102097). The funder played no role in study design, writing of this manuscript or the decision to submit for publication. The authors would like to acknowledge intellectual input from the ORBIT investigator team: Neil Thomas, Erin Michalak, Lesley Berk, Michael Berk, Steve Bowe, Sue Cotton, Lidia Engel, Sheri L Johnson, Steven Jones, Michael Kyrios, Sara Lapsley, Cathrine Mihalopoulos, and Tania Perich.

Conflicts of Interest

None declared.

References

1. Cuijpers P, van Straten A, Warmerdam L, van Rooy MJ. Recruiting participants for interventions to prevent the onset of depressive disorders: possible ways to increase participation rates. BMC Health Serv Res 2010 Jun 25;10:181 [FREE Full text] [doi: 10.1186/1472-6963-10-181] [Medline: 20579332]
2. Donker T, Blankers M, Hedman E, Lótsson B, Petrie K, Christensen H. Economic evaluations of Internet interventions for mental health: a systematic review. Psychol Med 2015 Dec;45(16):3357-3376. [doi: 10.1017/S0033291715001427] [Medline: 26235445]
3. Titov N, Dear B, Ali S, Zou J, Lorian C, Johnston L, et al. Clinical and cost-effectiveness of therapist-guided internet-delivered cognitive behavior therapy for older adults with symptoms of depression: a randomized controlled trial. Behav Ther 2015 Mar;46(2):193-205. [doi: 10.1016/j.beth.2014.09.008] [Medline: 25645168]
4. Lal S, Adair C. E-mental health: a rapid review of the literature. Psychiatr Serv 2014 Jan 01;65(1):24-32. [doi: 10.1176/appi.ps.201300009] [Medline: 24081188]
5. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
6. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014 Oct;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
7. Andrews G, Basu A, Cuijpers P, Craske M, McEvoy P, English C, et al. Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: An updated meta-analysis. J Anxiety Disord 2018 Apr;55:70-78 [FREE Full text] [doi: 10.1016/j.janxdis.2018.01.001] [Medline: 29422409]
8. Murray G, Leitan N, Berk M, Thomas N, Michalak E, Berk L, et al. Online mindfulness-based intervention for late-stage bipolar disorder: pilot evidence for feasibility and effectiveness. J Affect Disord 2015 Jun 01;178:46-51. [doi: 10.1016/j.jad.2015.02.024] [Medline: 25795535]
9. Fletcher K, Foley F, Thomas N, Michalak E, Berk L, Berk M, et al. Web-based intervention to improve quality of life in late stage bipolar disorder (ORBIT): randomised controlled trial protocol. BMC Psychiatry 2018 Jul 13;18(1):221 [FREE Full text] [doi: 10.1186/s12888-018-1805-9] [Medline: 30001704]
10. Susman JL. Improving outcomes in patients with bipolar disorder through establishing an effective treatment team. Prim Care Companion J Clin Psychiatry 2010;12(Suppl 1):30-34 [FREE Full text] [doi: 10.4088/PCC.9064su1c.05] [Medline: 20628504]
11. Bauer MS, Biswas K, Kilbourne AM. Enhancing multiyear guideline concordance for bipolar disorder through collaborative care. Am J Psychiatry 2009 Nov;166(11):1244-1250. [doi: 10.1176/appi.ajp.2009.09030342] [Medline: 19797436]
12. Leitan N, Michalak E, Berk L, Berk M, Murray G. Optimizing delivery of recovery-oriented online self-management strategies for bipolar disorder: a review. Bipolar Disord 2015 Mar;17(2):115-127. [doi: 10.1111/bdi.12258] [Medline: 25238632]
13. Christensen H, Petrie K. State of the e-mental health field in Australia: where are we now? Aust N Z J Psychiatry 2013 Feb;47(2):117-120. [doi: 10.1177/0004867412471439] [Medline: 23297367]
14. Case AD, Byrd R, Claggett E, DeVeaux S, Perkins R, Huang C, et al. Stakeholders' perspectives on community-based participatory research to enhance mental health services. Am J Community Psychol 2014 Dec;54(3-4):397-408. [doi: 10.1007/s10464-014-9677-8] [Medline: 25245601]
15. Thomas N, Farhall J, Foley F, Leitan ND, Villagonzalo KA, Ladd E, et al. Promoting Personal Recovery in People with Persisting Psychotic Disorders: Development and Pilot Study of a Novel Digital Intervention. Front Psychiatry 2016;7:196 [FREE Full text] [doi: 10.3389/fpsyt.2016.00196] [Medline: 28066271]
16. Thomas N, McLeod B, Jones N, Abbott J. Developing Internet interventions to target the individual impact of stigma in health conditions. Internet Interventions 2015 Sep;2(3):351-358. [doi: 10.1016/j.invent.2015.01.003]
17. Todd N, Solis-Trapala I, Jones S, Lobban F. An online randomised controlled trial to assess the feasibility, acceptability and potential effectiveness of 'Living with Bipolar': a web-based self-management intervention for bipolar disorder: trial design and protocol. Contemp Clin Trials 2012 Jul;33(4):679-688. [doi: 10.1016/j.cct.2012.02.011] [Medline: 22387150]
18. Wilhelmsen M, Lillevoll K, Risør MB, Høifødt R, Johansen M, Waterloo K, et al. Motivation to persist with internet-based cognitive behavioural treatment using blended care: a qualitative study. BMC Psychiatry 2013 Nov 07;13:296 [FREE Full text] [doi: 10.1186/1471-244X-13-296] [Medline: 24199672]

19. Murray G, Leitan N, Thomas N, Michalak E, Johnson S, Jones S, et al. Towards recovery-oriented psychosocial interventions for bipolar disorder: Quality of life outcomes, stage-sensitive treatments, and mindfulness mechanisms. Clin Psychol Rev 2017 Mar;52:148-163. [doi: 10.1016/j.cpr.2017.01.002] [Medline: 28129636]
20. Fives A, Rusell D, Kearns N, Lyons R, Eaton P, Canavan J, et al. The role of random allocation in randomized controlled trials: distinguishing selection bias from baseline imbalance. J MultiDiscipl Eval 2013;9(20):33-42 [FREE Full text]
21. Murray E. Web-based interventions for behavior change and self-management: potential, pitfalls, and progress. Med 2 0 2012;1(2):e3 [FREE Full text] [doi: 10.2196/med20.1741] [Medline: 25075231]
22. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
23. Donkin L, Christensen H, Naismith S, Neal B, Hickie I, 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]
24. Proudfoot J, Klein B, Barak A, Carlbring P, Cuijpers P, Lange A, et al. Establishing guidelines for executing and reporting Internet intervention research. Cogn Behav Ther 2011;40(2):82-97. [doi: 10.1080/16506073.2011.573807] [Medline: 25155812]
25. Doherty G, Coyle D, Sharry J. Engagement with online mental health interventions: an exploratory clinical study of a treatment for depression. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: ACM; 2012 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; May 2012; Austin, Texas p. 1421-1430 URL: https://dl.acm.org/citation.cfm?id=2208602
26. Pachankis JE, Lelutiu-Weinberger C, Golub SA, Parsons JT. Developing an online health intervention for young gay and bisexual men. AIDS Behav 2013 Nov;17(9):2986-2998 [FREE Full text] [doi: 10.1007/s10461-013-0499-8] [Medline: 23673791]
27. Graffigna G, Barello S, Bonanomi A, Lozza E. Measuring patient engagement: development and psychometric properties of the Patient Health Engagement (PHE) Scale. Front Psychol 2015;6:274 [FREE Full text] [doi: 10.3389/fpsyg.2015.00274] [Medline: 25870566]
28. Tetley A, Jinks M, Huband N, Howells K. A systematic review of measures of therapeutic engagement in psychosocial and psychological treatment. J Clin Psychol 2011 Sep;67(9):927-941. [doi: 10.1002/jclp.20811] [Medline: 21633956]
29. Kelders S, Kok R, Ossebaard H, 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]
30. 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 Dec;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
31. Murray G, Leitan N, Michalak E, Berk L, Berk M. Supporting self-management in bipolar disorder online: a review. 2015 Presented at: 17th Annual Conference of the International Society for Bipolar Disorders (ISBD 2015); June 2015; Toronto, Canada.
32. Todd N, Jones S, Hart A, Lobban F. A web-based self-management intervention for Bipolar Disorder 'living with bipolar': a feasibility randomised controlled trial. J Affect Disord 2014 Dec;169:21-29. [doi: 10.1016/j.jad.2014.07.027] [Medline: 25129531]
33. Lauder S, Chester A, Castle D, Dodd S, Berk L, Klein B, et al. Development of an online intervention for bipolar disorder. www.moodswings.net.au. Psychol Health Med 2013;18(2):155-165. [doi: 10.1080/13548506.2012.689840] [Medline: 22712771]
34. Todd NJ, Jones SH, Lobban FA. What do service users with bipolar disorder want from a web-based self-management intervention? A qualitative focus group study. Clin Psychol Psychother 2013;20(6):531-543. [doi: 10.1002/cpp.1804] [Medline: 22715161]
35. Naslund JA, Marsch LA, McHugo GJ, Bartels SJ. Emerging mHealth and eHealth interventions for serious mental illness: a review of the literature. J Ment Health 2015;24(5):321-332 [FREE Full text] [doi: 10.3109/09638237.2015.1019054] [Medline: 26017625]
36. Lauder S, Chester A, Castle D, Dodd S, Gliddon E, Berk L, et al. A randomized head to head trial of MoodSwings.net.au: an Internet based self-help program for bipolar disorder. J Affect Disord 2015 Jan 15;171:13-21. [doi: 10.1016/j.jad.2014.08.008] [Medline: 25282145]
37. Cosgrove V, Gliddon E, Berk L, Grimm D, Lauder S, Dodd S, et al. Online ethics: where will the interface of mental health and the internet lead us? Int J Bipolar Disord 2017 Dec;5(1):26 [FREE Full text] [doi: 10.1186/s40345-017-0095-3] [Medline: 28480484]
38. 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 [FREE Full text] [Medline: 9881538]
39. Pagliari C. Design and evaluation in eHealth: challenges and implications for an interdisciplinary field. J Med Internet Res 2007 May 27;9(2):e15 [FREE Full text] [doi: 10.2196/jmir.9.2.e15] [Medline: 17537718]

40. Ebert D, Van Daele T, Nordgreen T, Karekla M, Compare A, Zarbo C, et al. Internet- and Mobile-Based Psychological Interventions: Applications, Efficacy, and Potential for Improving Mental Health. European Psychologist 2018 Apr 12;23(2):167-187 [FREE Full text] [doi: 10.1027/1016-9040/a000318] [Medline: 28500020]
41. Thomas N, Foley F, Lindblom K, Lee S. Are people with severe mental illness ready for online interventions? Access and use of the Internet in Australian mental health service users. Australas Psychiatry 2017 Jun;25(3):257-261. [doi: 10.1177/1039856217689913] [Medline: 28139947]
42. Thomas N, Farhall J, Foley F, Rossell SL, Castle D, Ladd E, et al. Randomised controlled trial of a digitally assisted low intensity intervention to promote personal recovery in persisting psychosis: SMART-Therapy study protocol. BMC Psychiatry 2016 Dec 07;16(1):312 [FREE Full text] [doi: 10.1186/s12888-016-1024-1] [Medline: 27604363]
43. Harding C, Chung H. Behavioral health support and online peer communities: international experiences. Mhealth 2016;2:43 [FREE Full text] [doi: 10.21037/mhealth.2016.10.04] [Medline: 28293613]
44. Gliddon E, Lauder S, Berk L, Cosgrove V, Grimm D, Dodd S, et al. Evaluating discussion board engagement in the MoodSwings online self-help program for bipolar disorder: protocol for an observational prospective cohort study. BMC Psychiatry 2015 Oct 14;15:243 [FREE Full text] [doi: 10.1186/s12888-015-0630-7] [Medline: 26462799]
45. Cavanagh K, Millings A. (Inter)personal Computing: The Role of the Therapeutic Relationship in E-mental Health. J Contemp Psychother 2013 Jul 17;43(4):197-206. [doi: 10.1007/s10879-013-9242-z]

Abbreviations

BD: bipolar disorder
CAG: consumer advisory group
HREC: Human Research Ethics Committee
ORBIT: Online, Recovery-Oriented Bipolar Individualised Tool
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 31.05.18; peer-reviewed by D Meyer, J Oldenburg; comments to author 05.07.18; revised version received 25.07.18; accepted 17.08.18; published 24.10.18.

Please cite as:
Fletcher K, Foley F, Murray G
Web-Based Self-Management Programs for Bipolar Disorder: Insights From the Online, Recovery-Oriented Bipolar Individualised Tool Project
J Med Internet Res 2018;20(10):e11160
URL: http://www.jmir.org/2018/10/e11160/
doi:10.2196/11160
PMID:30355553

Viewpoint

Teledermatology: Comparison of Store-and-Forward Versus Live Interactive Video Conferencing

Titus Josef Brinker[1,2,3], MD; Achim Hekler[1]; Christof von Kalle[1], MD, PhD; Dirk Schadendorf[4], MD, PhD; Stefan Esser[4,5], MD; Carola Berking[6], MD; Martina T Zacher[6], MD; Wiebke Sondermann[4], MD; Niels Grabe[7], PhD; Theresa Steeb[6], MPH; Jochen Sven Utikal[8], MD; Lars E French[9], MD; Alexander H Enk[2], MD

[1]Department of Translational Oncology, National Center for Tumor Diseases, German Cancer Research Center, Heidelberg, Germany
[2]Department of Dermatology, University Hospital Heidelberg, University of Heidelberg, Heidelberg, Germany
[3]German Cancer Consortium, University of Heidelberg, Heidelberg, Germany
[4]Department of Dermatology, University Hospital Essen, University of Duisburg-Essen, Essen, Germany
[5]University Hospital Zurich, University of Zurich, Zurich, Switzerland
[6]Department of Dermatology and Allergology, University Medical Center Munich, Ludwig Maximilian University of Munich, Munich, Germany
[7]BioQuant, Hamamatsu Tissue Imaging and Analysis Center (TIGA), University of Heidelberg, Heidelberg, Germany
[8]German Cancer Research Center, Skin Cancer Unit, University of Heidelberg, Mannheim, Germany
[9]Department of Dermatology, University Hospital of Zurich, University of Zurich, Zurich, Switzerland

Corresponding Author:
Titus Josef Brinker, MD
Department of Translational Oncology
National Center for Tumor Diseases
German Cancer Research Center
Im Neuenheimer Feld 460
Heidelberg,
Germany
Phone: 49 151 7508 4347
Email: titus.brinker@nct-heidelberg.de

Abstract

A decreasing number of dermatologists and an increasing number of patients in Western countries have led to a relative lack of clinicians providing expert dermatologic care. This, in turn, has prolonged wait times for patients to be examined, putting them at risk. Store-and-forward teledermatology improves patient access to dermatologists through asynchronous consultations, reducing wait times to obtain a consultation. However, live video conferencing as a synchronous service is also frequently used by practitioners because it allows immediate interaction between patient and physician. This raises the question of which of the two approaches is superior in terms of quality of care and convenience. There are pros and cons for each in terms of technical requirements and features. This viewpoint compares the two techniques based on a literature review and a clinical perspective to help dermatologists assess the value of teledermatology and determine which techniques would be valuable in their practice.

(J Med Internet Res 2018;20(10):e11871) doi:10.2196/11871

KEYWORDS

teledermatology; live video conferencing; store-and-forward teledermatology; mobile phone; wait time; live interactive

Introduction

Background

Teledermatology, originating in 1995, was one of the first telemedicine services to be implemented, with other medical specialties following later [1-3]. Aging populations and a relative lack of dermatologists have prolonged wait times in the Western world, increasing the demand for new, more efficient strategies to render dermatologic care [4]. Mobile phones have overcome the image resolution limitations seen with older devices, opening a new field of mobile teledermatology, with two approaches in use. Store-and-forward (SAF) teledermatology allows transmission of images and text to a clinician for review. Live video conferencing (LVC), on the other hand, allows patient and physician to meet virtually at the same time using a webcam or mobile phone camera.

Determination of Diagnostic Accuracy

Introducing any new diagnostic method in patient care requires testing to ensure the diagnostic accuracy is at least comparable to the accepted standard. Evaluating diagnostic accuracy in dermatology is quite complex. A clinical diagnosis made by a specialist of a lesion as benign is accepted as the reference standard. When a biopsy is performed, however, the reference standard is clearly the histopathology result. Yet even among pathologists, there may be considerable discord in distinguishing between melanoma and benign melanocytic lesions. A review of 392 cases in 2010 in the United States revealed discordant results between pathologists in 14.3% of cases [5]. A 2016 study from another US center indicated discord in 114 of 588 cases (19.4%) [6]. Given the level of disagreement between histopathologists, including those specializing in dermatohistopathology, studies investigating the diagnostic accuracy of a new method, compared with current standards, must be interpreted cautiously [5-7]. Misdiagnosis obviously can have serious impact on patients, but it also complicates studies of newer diagnostic methods.

Diagnostic Accuracy of Teledermatology

A 2016 review systematically analyzed 21 studies, comparing teledermatology diagnoses using SAF or LVC with results of histopathology or, for nonexcised lesions, clinical diagnoses from face-to-face (FTF) encounters [8]. Overall, FTF diagnosis performed slightly better (67% to 85% agreement with the reference standard, Cohen kappa=0.90) compared with teledermatology (51% to 85% agreement, kappa=0.41–0.63) for the diagnosis of skin cancer. However, several studies have reported teledermatology is more accurate, in some cases even better than in FTF encounters, perhaps because of the improved resolution of mobile phone cameras [9,10].

In the case of skin cancer, timely management is crucial. A review of 3 studies by Finnance et al [8] reported significantly shorter wait times for melanoma patients assessed by SAF mobile phone technology [11,12] compared with conventional procedures. Patients who were referred using teledermatology triage systems tended to receive primary treatment at the first dermatology appointment and required fewer repeat visits [11,12]. We found no data for LVC on this aspect.

Diagnostic Accuracy With Mobile Phone Dermoscopes

While dermoscopic evaluation is the clinical practice gold standard for FTF visits and has been proven to increase diagnostic accuracy [13], teledermoscopy requires the patient to purchase a dermoscope to use with a mobile phone even though it may not necessarily be superior to teledermatology alone. In a landmark publication in 2011, Krömer et al [14] reported that teledermoscopy had a very high sensitivity and specificity for both malignant melanocytic lesions (sensitivity, 100%; specificity, 97% to 98%; n=6) and malignant nonmelanocytic lesions (sensitivity, 97%; specificity, 92% to 94%, n=58). There was no significant difference between the clinical and dermoscopic diagnoses based on histopathology as the reference standard. The authors reported that, in terms of detailed diagnoses, there were only 16 discordant diagnoses with teledermatology versus 22 with teledermoscopy [14]. A study by Senel et al [15] found that management plans based on teledermatology did not differ significantly from those developed in a FTF encounter, although the accuracy was significantly improved with a mobile phone dermoscope. Further study will help determine what, if any, value is added by teledermoscopy.

General Skin Conditions

A number of teledermatology studies have focused on any visible skin condition, including a large proportion of nonmalignant lesions that were not biopsied. In these studies, the reference standard was the clinical diagnosis in a FTF visit, so diagnostic concordance conclusions were limited. Overall, however, both original studies and reviews confirm improved teledermatology diagnostic accuracy, particularly because of improved digital image resolution. These investigators conclude that teledermatology now had a diagnostic accuracy comparable to that in a FTF encounter [9,16,17].

Discussion

Comparison of Store-and-Forward and Live Video Conferencing

Data on direct comparisons of SAF and LVC in terms of diagnostic accuracy remain scarce [8]. A 2017 study of 214 patients examined video image resolution, assessing several teledermatology formats for concordance with FTF examination as the reference standard. SAF and uncompressed video results were similar and were significantly better than lower resolution compressed video [18]. Uncompressed video may, therefore, close the resolution gap between LVC and SAF methods, although it requires the user to have a faster internet connection.

Comparison of Requirements for Store-and-Forward and Live Video Conferencing

SAF and LVC have different requirements which, in turn, affect their suitability for different patient subgroups and ultimate benefit in terms of care [16]. SAF has a number of advantages over LVC in terms of equipment and timing (Table 1), such that it may be preferable in more settings.

Interpretation of Requirements

SAF appears preferable for both the patient and dermatologist in terms of equipment and time requirements. This would particularly be the case in areas where a fast or stable internet connection is unavailable. SAF, therefore, would likely increase the number of patients for whom teledermatology is available. It might also attract more clinicians to engage in it, since it offers more flexibility than either routine care or LVC. Hook et al [19] noted a particular advantage of the anonymity available with SAF teledermatology, as patients with lesions in sensitive areas (eg, from sexually transmitted diseases) may not be willing to identify themselves to a doctor, potentially delaying diagnosis and treatment.

Table 1. Advantages of store-and-forward over live video conferencing in teledermatology.

Requirements	Store-and-forward	Live video conferencing
Availability of internet connection	Can be prepared without internet (ie, photos, history) and uploaded or downloaded at any time.	Simultaneous and continuous internet connection is required for both parties.
Speed of internet connection	Internet speed unimportant.	Slow internet speed may lessen diagnostic accuracy.
Appointment	No appointment necessary as evaluation is asynchronous.	Appointment required for synchronous evaluation.
Webcam or mobile phone camera	Useful but not required. Pictures may be on file or taken with any device.	Webcam or mobile phone camera required for both parties for entire session.
Personal identification	Anonymous access possible.	Identification required for face-to-face consultation.

Table 2. Features of store-and-forward and live video conferencing for teledermatology.

Features	Store-and-forward	Live video conferencing
Diagnostic accuracy	Higher for store-and-forward compared with that for low-resolution live video conferencing.[a]	Equally high if high-resolution live video conferencing is used.[a]
Physician-patient interaction	Usually low. Physician response may include asking for more information or other images by text.	Physician can directly ask the patient to perform certain tasks or show certain body regions.
Image resolution	Resolution of photographic images is usually higher.	Video images usually have a lower resolution.
3-dimensional view	Not possible for static images.	Live video feed allows clinician to view lesions from various angles.
Webcam or mobile phone camera	Useful but not required, as pictures may be taken with any device.	Webcam or mobile phone camera is required for both parties for entire session.
Free choice of location	Full flexibility.	Bound to locations with a fast internet connection and appropriate equipment.
Free choice of time	Full flexibility.	Bound to scheduled appointment.
Teledermoscopy	Often conducted.[b]	No published literature.
Wait time	Reduced.	No data.
Cost effectiveness	Higher.	Lower.
Security of data transmission	Most commonly transport layer security protocol (end-to-end encryption)	Most commonly transport layer security protocol (end-to-end encryption)

[a] Depends on setting, reference standard, and technology.

[b] Data on increased diagnostic accuracy inconclusive.

Comparison of Features of Store-and-Forward and Live Video Conferencing

Teledermatology method preferences may depend on the particular features (Table 2) desired by dermatologists and patients.

In regions where many households have no broadband internet connection, such as Germany and Switzerland, SAF is preferable over LVC because of higher diagnostic accuracy [18]. Inner cities offer a better availability of broadband internet connection than rural areas. In regard to an aging population, this aspect should be reinforced as the average age of the population in rural areas is higher which leads to an increased demand for accessible dermatologic care. In addition, SAF offers more time independence for both patient and clinician, letting older people take their time in setting up the technology before sending in their case. Interaction between patient and clinician about SAF images depends on the software used. In some cases, the dermatologist can chat with the patient to ask questions after reviewing the images. The fact that LVC allows synchronous interaction does not appear to increase its diagnostic accuracy over SAF, even with high-resolution LVC [18]. SAF image resolution is higher and can include teledermoscopy. In fact, a systematic review showed SAF did no harm or even improved accuracy in diagnosing skin cancer [20]. Patients at high risk of melanoma are reportedly very accepting of teledermoscopy [21]. While SAF imaging is not currently available in 3D, sensor-in-motion technology that transfers a stored high-definition video along with standard images may become available. Even static images are useful if they are taken from at least two different angles. From a cost effectiveness standpoint, both methods have been shown to reduce costs, but LVC has been found to be more expensive than an SAF approach due to the need for expensive video conference equipment and more consultation time [22]. With regard to patient comfort, 18% of patients reported feeling uncomfortable and 17% reported feeling embarrassed during LVC [23]. However, in some cases SAF patients reported dissatisfaction with the absence of a face-to-face office visit with a dermatologist and when being photographed by another person [24].

Conclusion

The authors regard SAF as the standard of care for teledermatology. It is well supported by evidence in the literature and available to more patients than LVC (in terms of both location and equipment), offers greater privacy, reduces wait times, improves access to care, and provides both clinicians and patients greater flexibility than traditional clinic visits. Nonetheless, teledermatology is complementary to and not a replacement of FTF clinical encounters.

Conflicts of Interest

TJB is the owner of a health technology company (Smart Health Heidelberg GmbH).

References

1. Nami N, Giannini E, Burroni M, Fimiani M, Rubegni P. Teledermatology: state-of-the-art and future perspectives. Expert Rev Dermatol 2014 Jan 10;7(1):1-3. [doi: 10.1586/edm.11.79]
2. Trettel A, Eissing L, Augustin M. Telemedicine in dermatology: findings and experiences worldwide—a systematic literature review. J Eur Acad Dermatol Venereol 2018 Feb;32(2):215-224. [doi: 10.1111/jdv.14341] [Medline: 28516492]
3. Fabbrocini G, De Vita V, Pastore F, D'Arco V, Mazzella C, Annunziata M, et al. Teledermatology: from prevention to diagnosis of nonmelanoma and melanoma skin cancer. Int J Telemed Appl 2011;2011:125762 [FREE Full text] [doi: 10.1155/2011/125762] [Medline: 21776252]
4. Coates SJ, Kvedar J, Granstein RD. Teledermatology: from historical perspective to emerging techniques of the modern era: part I: history, rationale, and current practice. J Am Acad Dermatol 2015 Apr;72(4):563-575. [doi: 10.1016/j.jaad.2014.07.061] [Medline: 25773407]
5. Shoo BA, Sagebiel RW, Kashani-Sabet M. Discordance in the histopathologic diagnosis of melanoma at a melanoma referral center. J Am Acad Dermatol 2010 May;62(5):751-756. [doi: 10.1016/j.jaad.2009.09.043] [Medline: 20303612]
6. Patrawala S, Maley A, Greskovich C, Stuart L, Parker D, Swerlick R, et al. Discordance of histopathologic parameters in cutaneous melanoma: clinical implications. J Am Acad Dermatol 2016 Jan;74(1):75-80. [doi: 10.1016/j.jaad.2015.09.008] [Medline: 26514601]
7. Heal C, Weedon D, Raasch B, Hill B, Buettner PG. Agreement between histological diagnosis of skin lesions by histopathologists and a dermato-histopathologist. Int J Dermatol 2009 Dec;48(12):1366-1369. [doi: 10.1111/j.1365-4632.2009.04185.x] [Medline: 19930494]
8. Finnane A, Dallest K, Janda M, Soyer HP. Teledermatology for the diagnosis and management of skin cancer: a systematic review. JAMA Dermatol 2017 Mar 01;153(3):319-327. [doi: 10.1001/jamadermatol.2016.4361] [Medline: 27926766]
9. Nami N, Massone C, Rubegni P, Cevenini G, Fimiani M, Hofmann-Wellenhof R. Concordance and time estimation of store-and-forward mobile teledermatology compared to classical face-to-face consultation. Acta Derm Venereol 2015 Jan;95(1):35-39 [FREE Full text] [doi: 10.2340/00015555-1876] [Medline: 24889827]
10. van der Heijden J, Thijssing L, Witkamp L, Spuls P, de Keizer NF. Accuracy and reliability of teledermatoscopy with images taken by general practitioners during everyday practice. J Telemed Telecare 2013 Sep;19(6):320-325. [doi: 10.1177/1357633X13503437] [Medline: 24163296]
11. Börve A, Dahlén GJ, Terstappen K, Johansson BE, Aldenbratt A, Danielsson M, et al. Smartphone teledermoscopy referrals: a novel process for improved triage of skin cancer patients. Acta Derm Venereol 2015 Feb;95(2):186-190 [FREE Full text] [doi: 10.2340/00015555-1906] [Medline: 24923283]
12. Morton C, Downie F, Auld S, Smith B, van der Pol M, Baughan P, et al. Community photo-triage for skin cancer referrals: an aid to service delivery. Clin Exp Dermatol 2011 Apr;36(3):248-254. [doi: 10.1111/j.1365-2230.2010.03960.x] [Medline: 21070338]
13. Salerni G, Terán T, Puig S, Malvehy J, Zalaudek I, Argenziano G, et al. Meta-analysis of digital dermoscopy follow-up of melanocytic skin lesions: a study on behalf of the International Dermoscopy Society. J Eur Acad Dermatol Venereol 2013 Jul;27(7):805-814. [doi: 10.1111/jdv.12032] [Medline: 23181611]
14. Kroemer S, Frühauf J, Campbell TM, Massone C, Schwantzer G, Soyer HP, et al. Mobile teledermatology for skin tumour screening: diagnostic accuracy of clinical and dermoscopic image tele-evaluation using cellular phones. Br J Dermatol 2011 May;164(5):973-979. [doi: 10.1111/j.1365-2133.2011.10208.x] [Medline: 21219286]
15. Senel E, Sabancılar E, Mansuroğlu C, Demir E. A preliminary study of the contribution of telemicroscopy to the diagnosis and management of skin tumours in teledermatology. J Telemed Telecare 2014 Jun;20(4):178-183. [doi: 10.1177/1357633X14533885] [Medline: 24786415]
16. Augustin M, Wimmer J, Biedermann T, Blaga R, Dierks C, Djamei V, et al. [Praxis der Teledermatologie]. J Dtsch Dermatol Ges 2018 Jul;16(Suppl 5):6-57. [doi: 10.1111/ddg.13512] [Medline: 29998512]
17. Warshaw E, Hillman Y, Greer N, Hagel E, MacDonald R, Rutks I, et al. Teledermatology for diagnosis and management of skin conditions: a systematic review. J Am Acad Dermatol 2011 Apr;64(4):759-772. [doi: 10.1016/j.jaad.2010.08.026] [Medline: 21036419]

18. Marchell R, Locatis C, Burges G, Maisiak R, Liu W, Ackerman M. Comparing high definition live interactive and store-and-forward consultations to in-person examinations. Telemed J E Health 2017 Dec;23(3):213-218 [FREE Full text] [doi: 10.1089/tmj.2016.0093] [Medline: 27705083]
19. Hook EW, Richey CM, Leone P, Bolan G, Spalding C, Henry K, et al. Delayed presentation to clinics for sexually transmitted diseases by symptomatic patients: a potential contributor to continuing STD morbidity. Sex Transm Dis 1997 Sep;24(8):443-448. [Medline: 9293606]
20. Bruce A, Mallow J, Theeke L. The use of teledermoscopy in the accurate identification of cancerous skin lesions in the adult population: a systematic review. J Telemed Telecare 2018 Feb;24(2):75-83. [doi: 10.1177/1357633X16686770] [Medline: 28056600]
21. Horsham C, Loescher LJ, Whiteman DC, Soyer HP, Janda M. Consumer acceptance of patient-performed mobile teledermoscopy for the early detection of melanoma. Br J Dermatol 2016 Dec;175(6):1301-1310. [doi: 10.1111/bjd.14630] [Medline: 27037999]
22. Loane M, Bloomer S, Corbett R, Eedy D, Hicks N, Lotery H, et al. A comparison of real-time and store-and-forward teledermatology: a cost-benefit study. Br J Dermatol 2000 Dec;143(6):1241-1247. [Medline: 11122028]
23. Whited JT. Teledermatology: current status and future directions. Am J Clin Dermatol 2001;2(2):59-64. [Medline: 11705304]
24. Williams TL, Esmail A, May CR, Griffiths CE, Shaw NT, Fitzgerald D, et al. Patient satisfaction with teledermatology is related to perceived quality of life. Br J Dermatol 2001 Dec;145(6):911-917. [Medline: 11899144]

Abbreviations

FTF: face-to-face
SAF: store-and-forward
LVC: live video conferencing

Edited by G Eysenbach; submitted 08.08.18; peer-reviewed by I Cosgarea, V De Vita; comments to author 29.08.18; revised version received 30.08.18; accepted 30.08.18; published 24.10.18.

Please cite as:
Brinker TJ, Hekler A, von Kalle C, Schadendorf D, Esser S, Berking C, Zacher MT, Sondermann W, Grabe N, Steeb T, Utikal JS, French LE, Enk AH
Teledermatology: Comparison of Store-and-Forward Versus Live Interactive Video Conferencing
J Med Internet Res 2018;20(10):e11871
URL: http://www.jmir.org/2018/10/e11871/
doi:10.2196/11871
PMID:30355564

Tutorial

A Gentle Introduction to the Comparison Between Null Hypothesis Testing and Bayesian Analysis: Reanalysis of Two Randomized Controlled Trials

Marcus Bendtsen[1], PhD

[1]Division of Community Medicine, Department of Medical and Health Sciences, Linköping University, Linköping, Sweden

Corresponding Author:
Marcus Bendtsen, PhD
Division of Community Medicine
Department of Medical and Health Sciences
Linköping University
Linköping, 58183
Sweden
Phone: 46 733140708
Email: marcus.bendtsen@liu.se

Abstract

The debate on the use and misuse of *P* values has risen and fallen throughout their almost century-long existence in scientific discovery. Over the past few years, the debate has again received front-page attention, particularly through the public reminder by the American Statistical Association on how *P* values should be used and interpreted. At the core of the issue lies a fault in the way that scientific evidence is dichotomized and research is subsequently reported, and this fault is exacerbated by researchers giving license to statistical models to do scientific inference. This paper highlights a different approach to handling the evidence collected during a randomized controlled trial, one that does not dichotomize, but rather reports the evidence collected. Through the use of a coin flipping experiment and reanalysis of real-world data, the traditional approach of testing null hypothesis significance is contrasted with a Bayesian approach. This paper is meant to be understood by those who rely on statistical models to draw conclusions from data, but are not statisticians and may therefore not be able to grasp the debate that is primarily led by statisticians.

(J Med Internet Res 2018;20(10):e10873) doi:10.2196/10873

KEYWORDS

null hypothesis testing; Bayesian analysis; randomized controlled trials; Bayes theorem; randomized controlled trials as topic

Introduction

Background

In response to a growing concern that claims of new discoveries as a result of scientific studies are becoming less and less credible, Benjamin et al [1] (signed by 71 authors) recommended that the threshold used to determine statistical significance should be reduced from the conventional .05 to .005. To do so, they claim, would immediately improve the reproducibility of scientific research in many fields. The authors acknowledge that any choice of threshold is arbitrary and that it incorporates a trade-off between false-positive and false-negative findings, yet they partially justified their choice of .005 by saying that it would reduce the false-positive rate to levels that they judge to be reasonable. In their concluding remarks, the authors pointed out that the proposed threshold should not be used to reject findings with *P* values between .005 and .05, but they should rather be labelled as suggestive evidence. Regarding this recommendation, Amrhein and Greenland [2] commented that, while this trichotomization may be better than the prevailing dichotomization into what is significant or not, it does not solve the issues of *P* hacking, selective reporting, and publication bias. Rather, the authors argued, it will only inflate these problems. Scientific conclusions should be based on multiple studies, and to allow for an unbiased and valid synthesis of the literature, all results must be published, regardless of *P* values. Furthermore, Amrhein and Greenland [2] pointed out that inference from a mathematical model cannot become “the truth” just because it passes some predefined threshold, and thus the authors suggested removing statistical significance completely.

Not only does the conventional null hypothesis testing using a threshold value of .05 constitute a requirement for publication, but as McShane et al [3] pointed out, it also constitutes a requirement for the results to be taken seriously. If the null hypothesis is not rejected, then researchers are stuck between two conditions in terms of conclusions, and are often far too eager to make a misinterpretation of no effect, since the null hypothesis was not rejected. Due to this, McShane et al [3] argued, considerations of the study design and quality of the data collected, prior and related evidence, plausibility of the mechanism that is investigated, novelty of the finding, real-world benefits and costs, etc, are only considered after *P* values have been checked, and if the threshold is broken, then little concern is given to these other factors. In this sense, whether or not statistical significance has been achieved has been given a superior standing over other equally important factors, and, since *P* values' main purpose has been to check for such significance, they too have been given elevated status. McShane et al [3] proposed putting *P* values on the same level as all other factors, thus abandoning statistical significance as an arbiter of truth, and treating *P* values as a continuous measure. The authors further argued that letting null hypothesis testing guide scientific discovery does not make sense, since the hypothesis tested is exactly no effect, which can never happen in an experimental setting and is in general very implausible (that an intervention has exactly no effect, whether it be positive or negative, is in most cases impossible). Thus, it is often forgotten that *P* values are calculated assuming a world in which the intervention has exactly no effect, but the probability of this world occurring is essentially zero. It should be emphasized that a *P* value is a mathematically correct and good answer to how likely a result is given a particular null hypothesis and may in some cases be a good enough approximation, but this in and of itself should not be a crucial factor for publication. McShane et al [3] support a holistic view of the evidence, in which all relevant factors are taken into consideration when interpreting statistical analyses, and this holistic view should also be adopted by journal editors and reviewers.

Voices have been raised over the past few years against the use and misuse of *P* values, perhaps most notably in a formal statement from the American Statistical Association clarifying widely agreed-upon principles underlying the proper use of *P* values [4], the banning of *P* values from the journal *Basic and Applied Social Psychology* [5], and Nuzzo's splendid summary in *Nature* [6]. McShane and Gal's [7] article is a fascinating read regarding the alarmingly widespread misinterpretation of *P* values and null hypothesis testing among both researchers who are not primarily statisticians and those who are. The ensuing discussion also gives an interesting insight into this problem and potential solutions [8-12].

Objective

This paper does not repeat the evidence put forward regarding the misinterpretation of *P* values, but instead contrasts the conventional null hypothesis and *P* value approach with that of a Bayesian analysis approach. The Bayesian approach taken is not in any sense novel, but has rather been proposed and used before; see Browne et al [13], Goodman and Sladky [14], Krushke [15], Morris et al [16], Spiegelhalter et al [17], and Wijeysundera et al [18]. However, as has been pointed out before [18], it is necessary to include nonstatisticians in the process of moving to a Bayesian approach. Therefore, this paper aims to inform those who routinely use null hypothesis testing and *P* values in the reporting of their research results, but who may not be responsible for running the analysis and may therefore find the discussion led by statisticians hard to grasp. Throughout, we attempt to give just enough understanding of the involved concepts so as to avoid too much technical detail, but at the same time we do not trivialize to the point where the discussion again becomes abstract. We begin by refreshing the reader's memory regarding probability distributions, since they play such a crucial role in statistical analyses, and then we use a coin flipping experiment to describe and contrast the conventional approach and the Bayesian approach. At this point, we turn our attention to real-world data, reanalyzing 2 randomized controlled trials. Finally, with a better understanding of the two approaches, we revisit the discussion outlined here.

Probability Distributions

As mentioned in the introduction, we do not attempt to offer an exhaustive discussion about the finer details of any mathematical aspects unless absolutely necessary. There is, however, no escaping the fact that it is necessary to understand, at least at a conceptual level, the notion of a probability distribution.

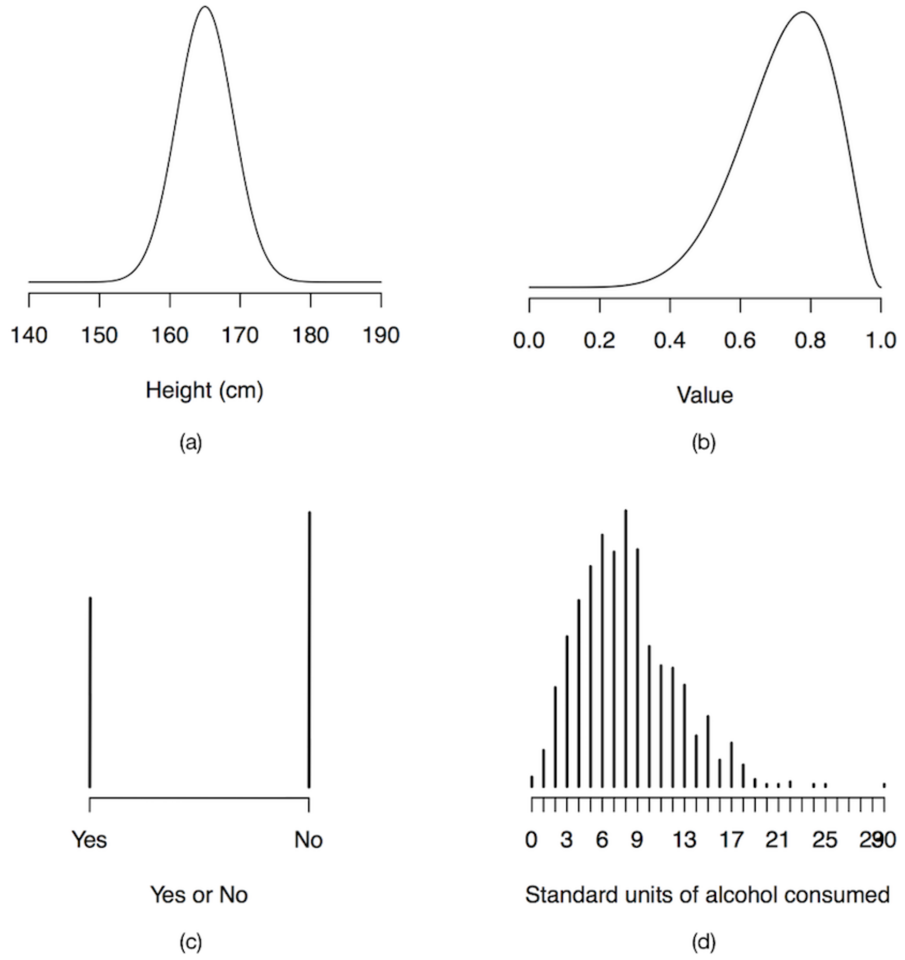
If we randomly pick a person from the general population, then we cannot, before we make our pick, possibly know their height. But we can do better than just saying that we know nothing about this person's height, since we do have an idea about people's heights in general. For instance, we know that the height cannot be negative and that it is unlikely to be more than 250 cm. Science requires us to reason in a systematic fashion, and for us to do so we need to express our knowledge about people's heights mathematically. Commonly this is done by assigning a probability distribution to our random person's height. A probability distribution is a purely mathematical construct that can tell us how likely different heights are relative to one another. So, it could tell us how likely it is that the person we pick will be between 160 and 180 cm tall, or how likely it is that the person will be taller than 150 cm. There are infinitely many probability distributions to pick from, and which one we use is our choice: we pick one that encodes our knowledge about people's heights. It should not be forgotten that probability distributions are mathematical constructs that help us create a systematic picture of the real world, but they make no claim to represent any truth about the real world.

For our purposes, we can think of probability distributions as shapes rather than mathematical equations. For instance, Figure 1 (part a) depicts a probability distribution for the height of a randomly picked person from the general population. We are here assuming that the distribution of heights can be represented using a *normal* distribution with a *mean* value of 165 cm and a *standard deviation* of 4 cm (this may not map perfectly to the real world; however, it is the choice that we have made). The mean and standard deviation are *parameters* of the normal distribution that tell us where we should center the distribution and how wide it is. Think of parameters as fine-tuning our choice of probability distribution—that is, we first picked the normal distribution and then we fined-tuned it using the parameters mean and standard deviation. Looking at Figure 1 (part a), we can see that values close to 165 cm are more likely than values further away, since the shape is higher around 165 cm. Figure 1 (part b) depicts a different distribution, known as the *beta* distribution. The beta distribution assigns probabilities only to values between 0 and 1; it would not be a good choice for modelling height, since we do not expect a person's height to be confined between 0 and 1, but the beta distribution can be used to model uncertainty about other problems. The beta has two parameters known as shape, so we can also fine-tune the beta distribution for our purposes.

In some cases, we have a finite number of outcomes. For instance, in a randomized controlled trial, we may have responses from participants to a yes-or-no question (eg, "Have you smoked any cigarettes the past week: yes or no"). In such cases, we can use a *Bernoulli* distribution that works over only two possible outcomes (a normal or beta distribution would not make sense here). A Bernoulli distribution has a parameter that we call q that tells us how likely it is that a participant will respond "no." Figure 1 (part c) depicts a Bernoulli distribution with the q parameter set to 0.6 (ie, we are encoding that there is a 60% chance that a participant answered "no"). As we can see, the shape is no longer a curve, but rather consists of bars that show how likely the outcomes are relative to one another. If we were investigating the number of whole standard units of alcohol consumed per week by a population, then we would have more than two possible outcomes, all greater than or equal to 0. In such cases, we could potentially use a *negative binomial* distribution, depicted in Figure 1 (part d), where we can see that there are more than two outcomes over which the distribution is defined.

The point to remember is that probability distributions allow us to encode uncertainty about quantities that we do not know the exact value of. For instance, if somebody asks what the height is of a randomly picked person off the street, we do not have to say "I do not know," but might instead answer "The height will follow a normal distribution with mean 165 cm and standard deviation 4 cm." There is a myriad of different probability distributions to pick from, and they all have different parameters that we can fine-tune to make sure that they encode our knowledge correctly. To understand most of this paper, we can think of probability distributions as shapes, just like the ones depicted in Figure 1 (a through d).

Figure 1. (a) A normal distribution with a mean value of 165 cm and standard deviation of 4 cm. (b) A beta distribution with shape parameters 8 and 3. (c) A Bernoulli distribution with q=0.6. (d) A negative binomial distribution with parameters 8 and 0.5 (failures and success probability).



Null Hypothesis Significance Testing and Bayesian Analysis of a Coin Flipping Experiment

We contrast the prevalent approach of null hypothesis significance testing (NHST) with a Bayesian analysis approach. For this comparison to be as simple as possible, in this section we use a classic experiment that we are all familiar with: flipping a coin and recording whether it lands heads or tails. Later, we compare the two approaches by reanalyzing 2 randomized controlled trials. However, to understand how the two approaches fundamentally differ, we begin by using a simple experiment and model.

Data and Model

Our experiment consists of flipping a coin 1000 times. We shall assume that the coin landed with heads up 540 times out of these 1000 flips. These are the data that we have collected: 540 heads and 460 tails. We would like to know whether the coin that we have used is fair—that is, whether the coin was manufactured in such a way that it is equally likely to get heads or tails when we flip it.

To encode and communicate the uncertainty about the outcome of flipping a coin, it is common to say that the outcome follows a Bernoulli distribution. We recall from the previous section on probability distributions that the Bernoulli distribution works over two possible outcomes (here we have heads or tails) and that it has a parameter q that in this experiment represents the probability of heads. We formally state our model as Equation 1:

$$\textit{coin flip} \sim \textit{Bernoulli}(q)$$
$$q = ?$$

The squiggly line should be read as “follows,” so that the model expresses the story “a coin flip follows a Bernoulli distribution with parameter q and the value of q is unknown.” Do not overanalyze Equation 1, as all it does is communicate to others that we believe that when we flip our coin there are two possible outcomes (Bernoulli) and that there is a probability q that our coin will land with heads up (but we do not know the value of q yet). It should be stressed that this is just a mathematical model of a coin flip, and there is nothing *true* about it. In fact, the model is actually wrong, since there is at least theoretically a third outcome, that the coin lands on edge standing straight up. A further infinite number of outcomes can be generated by considering the rotation of the coin.

We have our data (540 heads over 1000 flips) and our model in Equation 1, and our analysis should now revolve around the value of q. We therefore in the next two sections employ first an NHST approach and then a Bayesian approach to the analysis of q.

Null Hypothesis Significance Testing

When taking the NHST approach, we believe that there exists a fixed *population value* for q in Equation 1. In the coin flip experiment, it is easy to think of this population value as tied to some physical property of the coin. While one should avoid the word *true* when it comes to statistics, since all our inferences are based on a model that we have picked, we may think of this population value as the true value of q. In experiments involving a human study population, such as university students or office employees, the population value can be thought of as the value of q for the entire population. In most studies, we have only a sample of the entire population, in which case we cannot possibly know the population value for q. Note that it is not always clear what we mean by the population value, since study populations are often large to infinite in size, and sometimes the population is not very well defined (university students is a quite loosely defined group that changes from year to year). Nevertheless, the population value has a central role in the NHST approach.

Maximum Likelihood Estimator

We begin by considering the *maximum likelihood estimator* for q. This estimator is the value of q for which the likelihood of the data that we have collected is maximized. To decrypt what we mean by this, we can intuitively think of the maximum likelihood estimator as outlined in Textbox 1.

Returning to our original experiment, the maximum likelihood estimator for q would therefore be 540/1000=0.54 (recall that we had recorded 540 heads). It should, however, not come as a surprise that, if we went back and restarted the experiment and flipped the same coin 1000 times again, we would get a different outcome, for instance, 525 heads. This would then imply a different maximum likelihood estimate of 525/1000=0.525. In this way, we can see the maximum likelihood estimator as a proxy for the data that we have collected, a single number that summarizes information about the data with respect to the model.

As a side note, because of the rather simple model that we are employing (Equation 1), the maximum likelihood estimator was easy to calculate. It is, however, not always so, and for other models it may be necessary to apply optimization techniques to identify the maximum likelihood estimator. Most of us need not to worry about these details; we can assume that we can get a maximum likelihood estimator for most models.

Having calculated the maximum likelihood estimator, the next step is to consider a *sampling distribution*.

Textbox 1. Calculating the maximum likelihood estimator for q.

> Assume that we had recorded only 10 heads out of 1000 coin flips and that somebody suggests that the value of q should be 0.9, or a 90% probability of heads. Most of us would disagree and say that if q=0.9 then recording only 10 heads out of 1000 coin flips would be very unlikely. Another value might then be suggested, such as q=0.4, but we would still object, saying that 10 heads out of 1000 coin flips with a coin that is supposed to give 40% heads seems unlikely. So for which value of q would 10 heads in 1000 coin flips be most likely? It turns out that in this case it is trivial to calculate: 10/1000=0.01. So the value for q that makes 10 heads out of 1000 coin flips most likely is 1%, and this is the maximum likelihood estimator.

Sampling Distribution

From the discussion about maximum likelihood estimators, we concluded that, if we were to restart the coin flipping experiment, we could (even if we used the same coin) get a different number of heads. This would also then result in a different maximum likelihood estimator. Let us extend this line of thought and consider redoing the experiment thousands and thousands of times. What could we say about the maximum likelihood estimators that we would calculate for each one of these experiments? Just like we cannot know the exact height of a randomly picked person off the street before we actually pick and measure them, we cannot know which maximum likelihood estimator we will get next time we run the experiment. But this does not mean that we are totally unknowledgeable about the outcome: just like there is a distribution of heights, there is a distribution of maximum likelihood estimators. Theory tells us that this distribution is centered at the population value, and that it can be approximated by a normal distribution (at least when sample sizes are big enough). It is this distribution that is referred to as the sampling distribution. Each time we redo our coin flipping experiment, we get a maximum likelihood estimate that follows the sampling distribution (just like picking a person from the general population gives us a measurement of their height that follows the height distribution).

In our discussion about probability distributions, we mentioned that a normal distribution has two parameters: mean and standard deviation. The mean decides where the distribution is centered and the standard deviation decides how wide it is. We have established that the sampling distribution can be approximated by a normal distribution and that its mean (ie, its center) is the population value. Using our original data (540 heads over 1000 flips), we can use theoretical results to calculate an approximation of the standard deviation of the sampling distribution (often referred to as the standard *error*). In our case this value is approximately 0.0158.

Let us recapitulate. Given the data that we have collected (540 heads over 1000 flips), and the model that we have chosen (Equation 1), we can calculate a maximum likelihood estimate for q (540/1000=0.54), which follows a normal sampling distribution that is centered at the population value for q and has a standard deviation of 0.0158. Using this information, we can return to our original question: is the coin that we have flipped fair? To answer this, we turn to the practice of using hypothesis testing and *P* values.

Hypothesis Testing and P Values

We have previously stated that we wish to investigate whether the coin that we flipped was fair, and therefore our *null hypothesis* states that the population value for q is 0.5 (a q value of 0.5 means that there is a 50% probability of heads). If the null hypothesis fails to hold, we will instead accept an alternative hypothesis, which states that the population value for q is not 0.5 (ie, the coin is not fair).

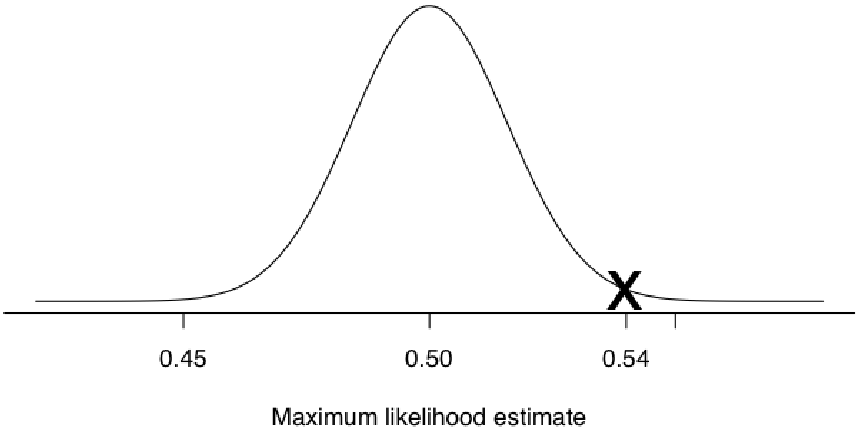
We now enter a hypothetical world in which we assume that the null hypothesis is true. This is a key concept: we are going to analyze our data in a world in which we know that the null hypothesis is true, and therefore the population value for q is known to be 0.5. Now recall that the sampling distribution that we defined previously was centered at the population value of q, but back then we did not know the population value. Now we know the mean and standard deviation of the sampling distribution and we can draw a representation of it. In Figure 2 we can see the sampling distribution for our maximum likelihood estimates (remember, just like people's heights have a distribution, so do maximum likelihood estimates). Additionally, in Figure 2 we have marked the maximum likelihood estimate that we calculated using the data that we collected in our experiment (540 heads over 1000 flips). Our maximum likelihood estimate is quite far out to the right, and such a value seems unlikely under this specific sampling distribution (the curve is low). Now recall that we said that the maximum likelihood estimate is a proxy for our data, a summary that we can use instead of the 1000 flips. This therefore tells us that the data that we have collected are quite unlikely, given that the population value is 0.5. But how unlikely? Enter *P* values. You sometimes hear people explain *P* values as "the probability of seeing these data or more extreme." It is sometimes hard to understand what is meant by more extreme data. What they are actually trying to say is "seeing this maximum likelihood estimate or higher" (or lower depending on which side of the center we are looking at).

Because we have approximated the sampling distribution using a normal distribution, it is easy to calculate the probability of a maximum likelihood estimator of 0.54 or more extreme given a mean value of 0.5 and a standard deviation of 0.0158. It turns out that this probability is approximately 0.0057, and we must multiply this value by 2 because we wish to do 2-sided tests (this has to do with the fact that we arbitrarily decided to do our calculations based on heads rather than tails). Therefore, our final *P* value is .0114.

Since this *P* value is less than the conventional threshold of .05, we say that the data that we have collected are so unlikely given the null hypothesis that we reject the null hypothesis and accept the alternative hypothesis. This is referred to as statistical significance. However, given the .005 threshold proposed by Benjamin et al [1], we cannot reject the null hypothesis and we would therefore not be able to say anything about the fairness of the coin.

To summarize, we enter a hypothetical world in which our null hypothesis is true, and if the data that we have collected seem unlikely or absurd in this world, then we reject the hypothesized world. But it does not say much about which world is the *true* world—that is, it does not say much about the population value. To narrow in on the population value, it is common to also report confidence intervals, which we turn to next.

Figure 2. Sampling distribution of *q* under the null hypothesis that the population value is 0.5. The X marks the maximum likelihood estimate of the coin flip experiment (540 heads over 1000 flips).



Confidence Intervals

Our maximum likelihood estimate is only 1 draw from the sampling distribution, so it does not tell us what the population value of *q* is. Recall that we are assuming that the population value of *q* is a fixed value, something that represents the entire population. One way of informing us about the location of the population value is to create what is known as confidence intervals.

Using a threshold of .05, we have already concluded that we will reject the null hypothesis that q=0.5, since the *P* value (.0114) was less than this threshold. We could increase our null hypothesis a bit, say to 0.501 rather than 0.5, and redo our hypothesis test as before. We would get a new *P* value of .0136, which would also lead to a rejection at the .05 threshold. But if we continue to increase the value of our null hypothesis, we would end up with a hypothesis that we cannot reject. This value is the lower limit of the confidence interval. Likewise, we can start from above our maximum likelihood estimate of 0.54 and find the largest value for our null hypothesis that cannot be rejected. The lowest and highest values that cannot be rejected are the confidence limits, and any hypothesis between these 2 limits cannot be rejected using the data that we have collected at the .05 threshold. In our coin flipping experiment, these limits are 0.509 and 0.571; thus, no hypotheses between these 2 values could be rejected given our data (540 heads over 1000 flips). Because we have chosen a threshold value of .05, these confidence intervals are known as 95% confidence intervals.

It would be nice if we could say that the population value of *q* lies within these 2 limits with 95% probability. But we cannot do so, unfortunately. Recall that if we could go back in time and redo the experiment, we would get a different maximum likelihood estimate; this means that we would also get a different set of confidence limits. What we can say, although it is very cumbersome, is that out of all the 95% confidence intervals that would be created by redoing the experiment, the population value for *q* will lie within them in 95% of the cases. If this sounds confusing, then you are in good company; most researchers tend to forget or misunderstand this.

Summary

This ends our introduction to the NHST approach. While we have attempted a high-level overview, we have nevertheless covered some central concepts that are necessary to keep in mind when applying this approach:

- The population value is a fixed value that we want to investigate.
- We collect data and compute maximum likelihood estimates for our model's parameters.
- We construct a sampling distribution (a distribution over maximum likelihood estimates).
- We hypothesize a population value, entering a world in which we assume that we know its true value.
- If, in the hypothesized world, the data are unlikely given some threshold, then we reject the null hypothesis—that is, we reject this world.
- We create confidence intervals, which tell us which hypotheses we cannot reject, and enable us to say something about the location of the population value (although this information might be very vague).

Bayesian

We have seen how the NHST approach focuses on understanding how likely the data gathered are given a sampling distribution and different hypothesized population values of *q*. The outcome of the analysis is information about which hypotheses we can and cannot refute given a predefined threshold. The Bayesian approach, however, asks the more direct question "How probable is every value of *q*?" There are an infinite number of *q* 's that we could pick, and the Bayesian wants to know how probable each one of them is, given the data that we have. The Bayesian approach does not rely on repeated experimentation to create a sampling distribution, but rather looks only at the probability for every *q* given the data that we have collected. What we receive by requesting this information is not a single value, such as the maximum likelihood estimate, but an entire distribution over all possible values of *q*.

The Bayesian philosophy is to begin with a belief about the quantity of interest (in our case, *q*), and then look at the data that have been collected and revise one's belief in light of the data. This is why words such as *updating* or *learning* are often used to describe the Bayesian approach, as we update our beliefs given the data, or alternatively learn something new from the data. To make this philosophy more formal, we rely on three concepts: prior distributions, data likelihood, and posterior distributions. We discuss these three in order, using the same coin flipping experiment as before.

Prior Distributions

When a quantity is unknown to us, such as *q* has been when we have been flipping coins, the Bayesian approach is to assign to this quantity a prior distribution. This prior distribution encodes our uncertainty about *q* before we analyze the data that we have collected. The keyword here is *before* (ie, prior). Just like the outcome of flipping a coin is unknown to us, so is the value of *q*. Our solution for describing the uncertainty about the coin flip was to say that it follows a Bernoulli distribution, and our solution for describing our uncertainty about *q* is to say that it also follows some distribution. There is a bit of harmony here, as we are not treating unknown quantities differently: as soon as the value of something is unknown to us, we say that it follows a distribution, regardless of whether it is data or parameters.

Recall that we have at our disposal many distributions that we can use to describe uncertainty: we have already encountered the normal, beta, Bernoulli, and negative binomial distributions. We also have the option of saying that we think that each value of *q* is equally likely before we analyze the data: we then say that *q* follows a uniform distribution. This is sometimes referred to as a *flat* prior, since the shape of the probability distribution is a flat line. Figure 3 (parts a through part c) presents three examples of different priors that we could choose: Figure 3 (part a) depicts a flat prior—that is, it assigns the same probability to every possible value of *q*. Figure 3 (part b) depicts a prior that says that we believe the coin to be fair before we start flipping it, so we assign more probability to *q* values around 0.5, but we are still assigning quite a bit of probability to all other values (the shape is wide). Finally, Figure 3 (part c) says that we believe the coin to be biased, assigning almost all probability to *q* values around 0.75 (the shape is very narrow).

When starting out with Bayesian analysis, it may seem like one would always want to pick a flat prior, like the one depicted in Figure 3 (part a). At first glance, this might seem like an objective choice, as there is no bias toward any specific value, and the NHST approach essentially takes this stance. However, this is not as objective as one might first think, and we shall return to this point in our discussion. Sometimes it may be beneficial to pick priors that enable analysis, for instance, if the number of potential participants in a study is very low, expert information may be encoded into the prior allowing for the analysis to still output useful results; please see Goodman and Sladky [14] and Morris et al [16] for examples. Another case is when we have many covariates to choose from, but we wish to include only the relevant ones for the outcome in our model [19,20].

If we decide to use a flat prior for our coin flipping experiment, then we extend our model to express that before we collect any data we believe all values for *q* to be equally likely (Equation 2):

$$\text{coin flip} \sim \text{Bernoulli}(q)$$
$$q \sim \text{uniform}(0,1)$$

The equation now reads “We believe that coin flips follow a Bernoulli distribution and that the probability of heads is *q*. We also believe that *q* is equally likely to take on any value between 0 and 1.” Recall that only values between 0 and 1 make sense for *q*, since it represents the probability for heads, so we cannot have negative probabilities, nor probabilities above 1. Compare this with our original model in Equation 1, where we said that *q* was completely unknown; using priors forces us to be more specific and explicit about what we mean when we say that something is unknown.

That is all we need to say about priors for the moment. They make sure that we express the uncertainty about all unknown values up front before we start the analysis.

Data Likelihood

Akin to what we were calculating before, during the NHST discussion, the data likelihood tells us how remarkable the data that we have collected are given different values of *q*. If we propose that *q*=0.5, then we can calculate the probability that we would collect 540 heads over 1000 flips with this proposed value. We can make this calculation because we have chosen a model for our experiment (we chose a Bernoulli distribution); if we had no model, then we could not make any of these calculations. Intuitively, we would expect that if we had instead proposed that *q*=0.1, then the data should be less likely than when *q*=0.5, since we have collected 540 heads over 1000 flips.

Figure 3. (a) Uniform prior distribution (flat prior). (b) A prior distribution that encodes that fair coins are more likely. (c) A prior distribution that encodes that biased coins are more likely.

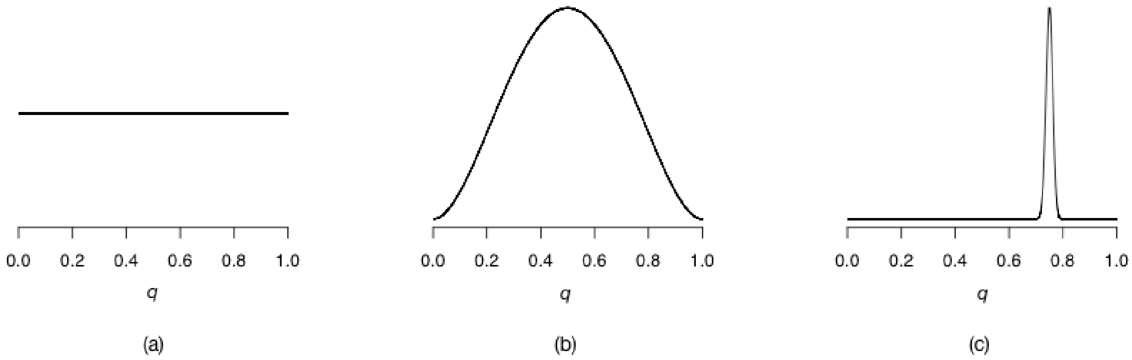
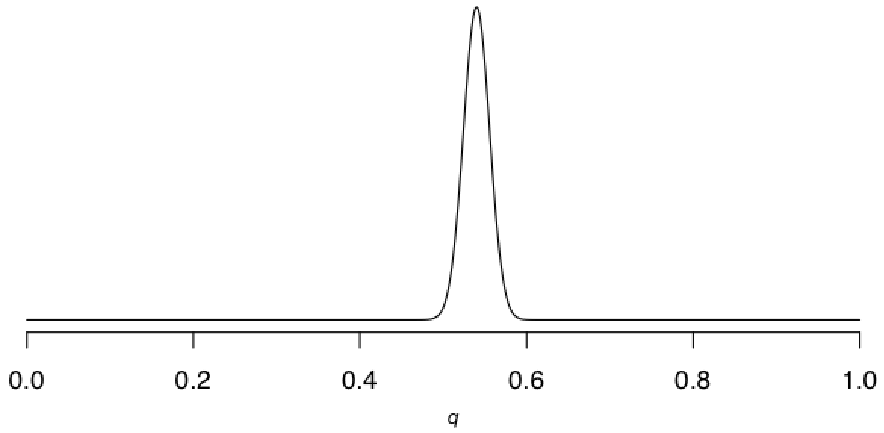


Figure 4. Data likelihood for different values of *q* in the coin flipping experiment.



If we continued this reasoning for every possible value of *q*, then we could draw a shape that tells us how likely the data that we have collected are for different values of *q*. Such a shape is drawn in Figure 4 for our coin flipping experiment. As we can see, this follows our intuition that the data are more likely given values for *q* around 0.5 compared with values around 0.1.

The shape in Figure 4 can be thought of as the data likelihood given our model. It tells us how remarkable the data that we have collected are given different values for *q*.

Posterior Distributions

The prior distribution encodes what we believe about *q* before we take into consideration any data, and the data likelihood tells us how remarkable the data that we have collected are given different values of *q*. But what we really care about is what we believe about *q after* we have taken into consideration the data. This is encoded in the *posterior* distribution, and it is the posterior distribution that is the answer to the Bayesian question "How probable is every value of *q*?"

The posterior distribution is a distribution just like all the others we have seen in this paper. It is calculated using Bayes' theorem. This theorem is a consequence of basic probability theory and named after famous statistician Reverend Thomas Bayes. Equation 3 is the simplified version. The theorem states that the posterior distribution can be computed by multiplying the data likelihood by the prior distribution.

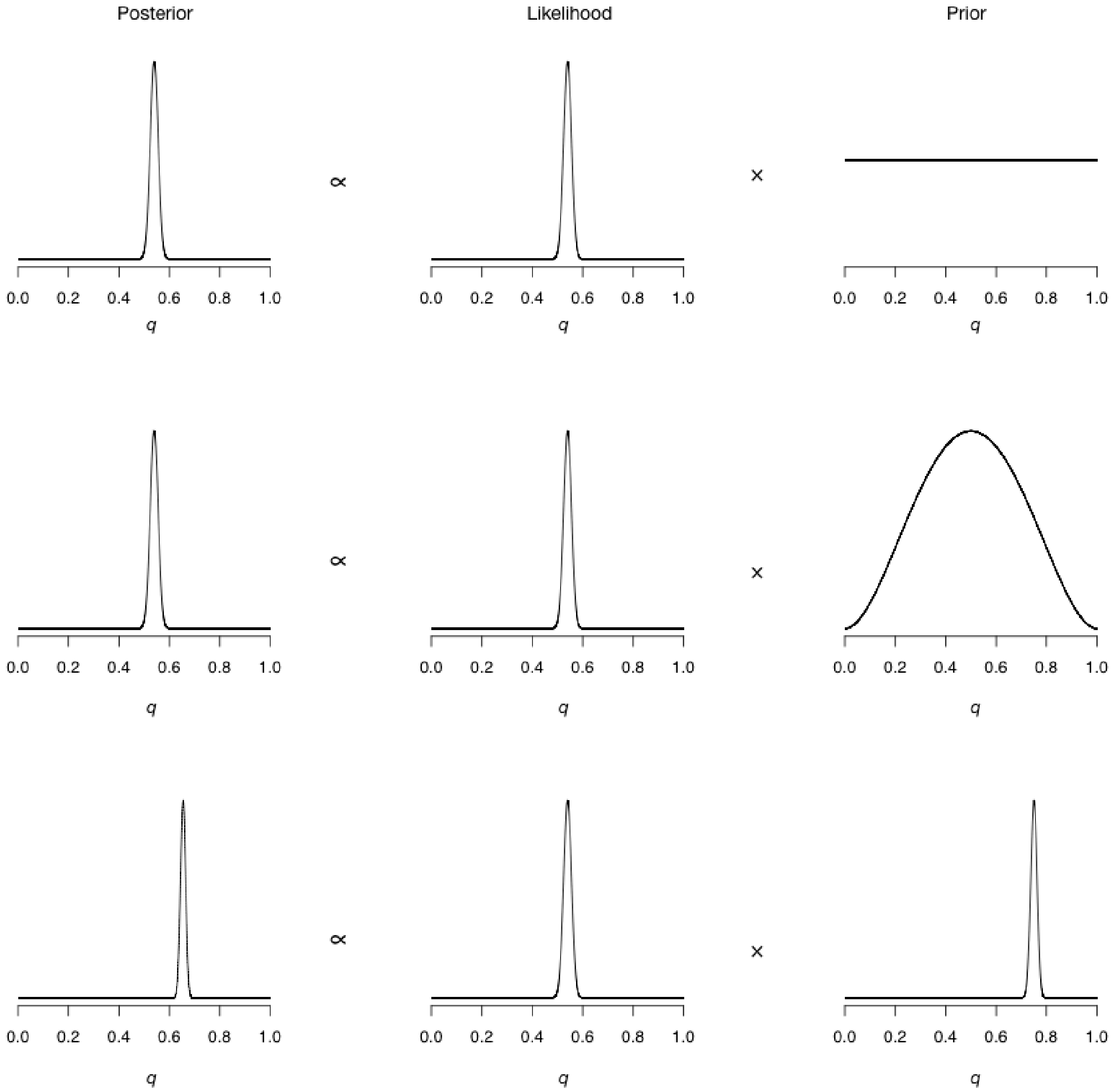
$$\textit{posterior} \propto \textit{likelihood} \times \textit{prior}$$

Rather than discussing this in terms of numbers, let us instead do this graphically, as we have been thinking of distributions as shapes rather than as equations. What we will be doing is essentially multiplying the priors that we depicted in Figure 3 a through 3c by the data likelihood depicted in Figure 4. In Figure 5, we can see Bayes' theorem in action for our coin flipping experiment. In each row we have a single use of the theorem. The top row shows us the result when using a flat prior, which is multiplied by the data likelihood to get a posterior. The second row shows us the use of the theorem with a prior that assigns more probability of the coin being fair, but does still allow for the entire range of possible *q* values (sometimes known as a weakly informative prior). The third and final row shows us the use of the theorem when we have a prior that very strongly believes that the coin is biased, using a very narrow prior around the value of 0.75 (note that this prior does not say that it is impossible that *q* can be 0.2, for instance; it just assigns a very small prior to this value of *q*). It is the column marked *Posterior* that is of interest at the moment. As we can see, the first 2 rows seem to have the same posterior: *q* values between 0.5 and 0.6 seem to be most probable according to these 2 rows. This is not a mistake; a common theme in Bayesian analysis is that once we have enough data the prior gets overwhelmed by the sheer amount of data. The prior that we picked for row number 2 assigned enough probability to all values of *q* that the data could easily overwhelm it, but not so for row number 3. In row number 3, we can see that the posterior distribution is shifted to the right; here values above 0.6 and less than 0.7 are more probable. The prior in the third row so strongly believed that the coin was biased that the data could not overwhelm it; thus, the entire posterior distribution is shifted toward the prior.

What we are saying is that the posterior probability of a value of *q* should take into consideration how likely this value was before we collected the data (the prior) but also how remarkable the data that we collected are under this value of *q*. So, for instance, if we were to collect 540 heads over 1000 flips and propose a *q* value of 0.01 (ie, a 1% chance of heads), then the posterior distribution for q=0.01 would be very low, since collecting 540 heads over 1000 flips when the probability of heads is only 1% is very unlikely. But proposing values around q=0.5 and q=0.6 should generate higher posterior probabilities, since 540 heads over 1000 flips is a lot more likely for such values.

We need not worry about the details of exactly how these calculations are done, but remember that Bayes' theorem is remarkably simple: the posterior is computed by multiplying the data likelihood by the prior distribution. Also note that the output of the Bayesian analysis is the posterior distribution—that is, a distribution over the parameter of interest (in this case *q*) after we have taken into consideration the data that we collected.

Figure 5. Three examples of the use of Bayes' theorem for the coin flipping experiment. Top row: posterior distribution when using a uniform prior distribution (flat prior). Middle row: posterior distribution when using a weak prior distribution that makes fair coins more likely. Bottom row: posterior distribution when using a strong prior distribution that makes biased coins more likely.

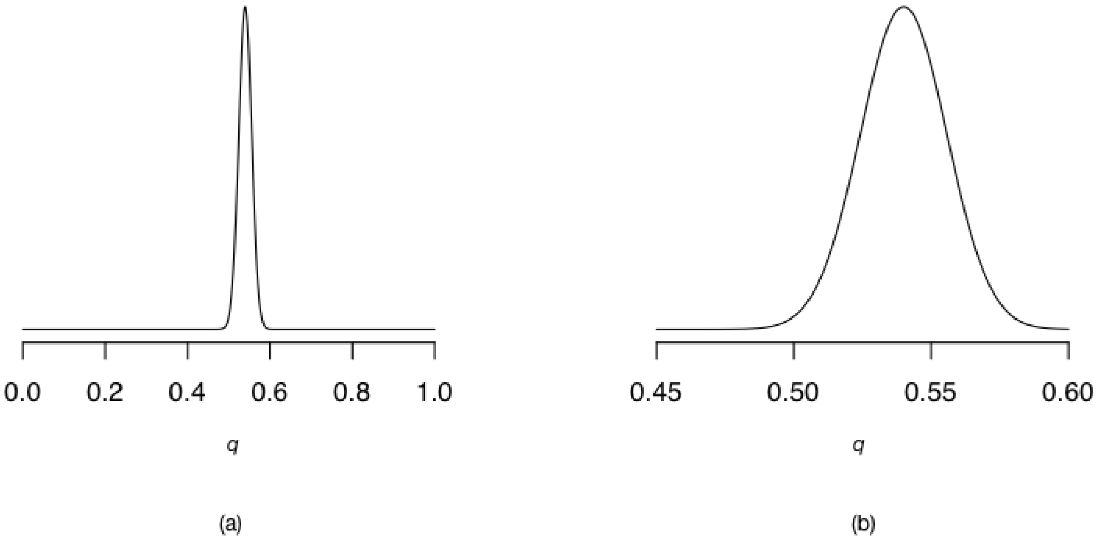


Analysis of the Coin Flipping Experiment

Figure 6 (part a) depicts the posterior distribution over *q* for our coin flipping experiment using a flat prior (this is the same as the top row in Figure 5). We have zoomed in on *q* values between 0.45 and 0.6 in Figure 6 (part b). What does the posterior distribution tell us? Just by looking at it, we can see that it is quite unlikely that *q*=0.2. However, it is important to note that we are not ruling out this case; it is still entirely possible that *q*=0.2, but given the coin flips that we have made it is logically less likely that *q*=0.2 compared with, say, 0.5. It seems that the most likely value of *q* relative to all others is around 0.54 (see the zoomed-in distribution in Figure 6, part b). It is, however, crucial to note that the result of the Bayesian analysis is not a single value such as 0.2, 0.5, 0.54, or 0.6, but rather the entire posterior distribution over *q*.

Once we have a posterior distribution over our parameter *q*, we can ask scientific questions about how probable different values of *q* are. We initially stated that we wished to investigate whether the coin was fair or not. A coin that is biased to resulting in more heads than tails would imply a *q* value greater than 0.5 (ie, there is a greater than 50% chance of heads), so we may ask "What is the probability that *q* is greater than 0.5?" The answer is given by the posterior distribution, and in this case it is approximately 99%. To see this, look at Figure 6 (part b) again and color the entire area underneath the curve above 0.5. As you can see, the area that you have colored far outweighs the area you have not. The story that we tell is, therefore, that "We flipped a coin 1000 times and 540 times it landed heads. There is a 99% probability that the coin is biased toward showing more heads than tails."

Figure 6. (a) Posterior distribution of q after observing 540 heads given 1000 flips ($0 \leq q \leq 1$). (b) Posterior distribution of q after observing 540 heads given 1000 flips ($0.45 \leq q \leq 0.6$).



In this case, it is hard to argue against the coin being biased because there was a 99% probability of it being so, but what is the conclusion if the probability was 60%? In the real-world data analysis that we will conduct, we shall encounter such a case and we shall therefore defer this discussion. Essentially, it ties into what McShane et al [3] referenced as neglected factors; that is, what are the real-world costs and benefits of the finding, how novel is this finding, given previous studies what does this finding tell us, etc. Definitive dichotomous conclusions belong to the NHST approach, not the Bayesian.

Summary

The Bayesian approach begins by assigning prior probability distributions to unknown quantities, extending our models to also encode uncertainty about the parameters. Using the likelihood of the data, the prior is updated using Bayes' theorem, resulting in a posterior distribution. The posterior distribution encodes the uncertainty about the model's parameters after we have taken the data into consideration.

We will now leave the fictitious coin flipping experiment that we have been treating here and instead focus on real-world data collected during randomized controlled trials. We will defer any contrasting between the NHST approach and the Bayesian approach described here to the general discussion section.

Analysis of Real-World Data

So far we have been using a rather trivial coin flipping example to illustrate the differences between the NHST and the Bayesian approaches. In this section, we instead look at data that were collected during 2 randomized controlled trials and complete a Bayesian analysis of the 2 trials in order to compare with the NHST analyses that have been published previously [21,22]. We shall look at the evaluation of a smoking cessation program and an alcohol consumption reduction program, both targeted at university students in Sweden and consisting of text messages sent to participants' mobile phones. We shall not delve into the details of the interventions, but will rather refer to them as the NEXit (for smoking) and AMADEUS (for alcohol) trials.

We begin by analyzing the NEXit trial: first, we describe the statistical model; second, we account for the NHST analysis already conducted; third, we conduct the new Bayesian analysis; and fourth, we discuss the outcome. We shall follow the same structure for the AMADEUS trial.

NEXit Trial

The NEXit trial was a single-blind, 2-arm, randomized controlled trial conducted between October 2014 and April 2015. Participants were daily or weekly smokers willing to set a quit date within 1 month of enrollment. Almost all college and university students in Sweden were contacted via email and invited to participate. Willing participants who fulfilled the inclusion criteria were randomly allocated to 2 groups: an intervention group that received the novel intervention and a control group that were asked to quit smoking on their own. The primary outcome measure was prolonged abstinence, defined as not having smoked more than 5 cigarettes during the past 8 weeks, and a 4-week point prevalence of complete smoking cessation (ie, no cigarettes smoked during the past 4 weeks). We shall not reanalyze any secondary outcomes.

Statistical Model

Both primary outcome measures in the NEXit trial were binary: participants responded either yes or no to the questions regarding prolonged abstinence and point prevalence. Just like in the coin flipping experiment, we are faced with two possible outcomes, and we do not know which outcome we will get if we randomly pick a NEXit participant. To reason systematically, we can say that the primary outcome measures in the NEXit trial follow a Bernoulli distribution with parameter q, where q represents the probability of a participant responding that they have not smoked (we treat each outcome measure separately). However, we would like to go a bit further and define a model that allows for different q values depending on whether a participant belongs to the control group or the intervention group, allowing us to contrast the difference between these q values. In a sense we wish to find 2 coins, 1 for each group, and compare whether one coin is more or less biased than the other.

The canonical way of modelling the narrative just given is to use what is known as logistic regression. We will avoid delving deeper into the details of this model, since the analysis here can be understood without them. What is important to note is that the quantity that is normally investigated is the *odds ratio* between the intervention group and the control group. The odds ratio is the odds of not smoking in the intervention group divided by the odds of not smoking in the control group. This quantity is convenient because it tells us by how much we should multiply the odds in the control group to get the odds in the intervention group. Thus, if the odds ratio is 1, then there is no effect, since you would take the control group's odds and multiply by 1, which gives the same result. If the odds ratio is greater than 1, for instance 2, then the intervention group has twice the odds of the control group of not smoking.

Do not overthink this. Before, we had a parameter q that described the probability of heads, and this was the parameter that we wished to investigate. Now we have the odds ratio, which is the quantity that we wish to investigate because we are comparing 2 coins. It is still just an unknown quantity that we wish to learn more about.

We begin by accounting for the original analysis that was done for the NEXit trial using the NHST approach, and then we shall account for a new Bayesian analysis of the data.

Null Hypothesis Significance Testing of the NEXit Trial

Of the 1590 participants randomly allocated into the NEXit trial, 1502 responded to follow-up regarding primary outcomes. Table 1 gives the maximum likelihood estimates for the odds ratios determined using logistic regression for the two primary outcome measures: the 95% confidence intervals and *P* values. Before the analysis, the researchers decided to perform 2-tailed tests at the .05 threshold. As Table 1 shows, the null hypothesis that the odds ratio is 1 (ie, no effect) was rejected (*P* values are <.05). Now recall that in the NHST approach, we use the maximum likelihood to *estimate* the fixed population odds ratio, and that the confidence interval should be interpreted such that the true population odds ratio lies within these limits for 95% of all the 95% confidence intervals that could be created if we were to redo the NEXit trial.

Bayesian Analysis of the NEXit Trial

The Bayesian approach begins by assigning prior probabilities to unknown quantities. We used flat priors for all unknown quantities, assigning equal probability to all values before seeing any data. This actually goes against our general recommendation, but we stick to flat priors so that we can defer any discussion about nonflat priors. Using Bayes' theorem, we computed posterior distributions over the unknown quantities and then use these posterior distributions to answer questions about the quantity of interest. In this case, we care about the odds ratio comparing the intervention group with the control group.

Figure 7 (part a) depicts the posterior distribution of the odds ratio of prolonged abstinence when comparing intervention versus control. Figure 7 (part b) similarly depicts the posterior distribution of the odds ratio of point prevalence when comparing intervention versus control.

The statistical model has done the statistical inference, and now it is up to the researcher to do the scientific inference. We have two outcome measures, which we have analyzed in terms of odds ratios. If the odds ratio is 1, then the intervention has no effect; if it is less than 1, then it has a negative effect; and if it is greater than 1, than it has a positive effect. We may therefore set up a series of questions to support our decision-making process. What is the probability that the odds ratios are greater than 1.0, 1.5, 2.0, and 2.5? The answers to these questions are given by the posterior distributions (this is why the outcome of a Bayesian analysis is the full distribution and not just a single value; we want to use the entire distribution to make a scientific inference). Table 2 summarizes the answers to these questions. As we can see, the posterior distribution tells us that it is very likely that the intervention had a positive effect on both prolonged and point prevalence outcome measures, since the posterior assigns more than 99% to these cases. It also seems more likely than not that the odds ratios for these outcomes were greater than 1.5. The odds ratio for prolonged abstinence is further more likely to be above 2.0 and, while the probability is severely lower at the 2.5 odds ratio, there is still 7.05% probability that the odds ratio is greater than 2.5.

The NEXit intervention is a fully automated intervention that does not require any interaction from health professionals. It is therefore cheap to offer and scales to large populations instantly. Participants are not put at harm and can stop the intervention at any time. It seems justifiable to offer the intervention to university students who want to quit smoking, given what the posterior distributions regarding prolonged and point prevalence abstinence tell us about the effect of NEXit. These posterior probabilities are of course calculated using a mathematical model that may or may not be a good approximation of the real world, so there is no escaping that one must assess the model chosen along with other factors. While we would like to confirm these results, and good research practice dictates that we should not blindly trust the results of a single study, if we assume that these are the only data available to us then the justification stands.

Table 1. Original analysis of the NEXit trial. Odds ratios compare intervention with control, given by logistic regression.

Outcome	Odds ratio	95% CI	*P* value[a]
Prolonged abstinence	2.05	1.58-2.66	≤.001
Point prevalence	1.57	1.19-2.05	.001

[a] 2-tailed.

Figure 7. (a) NEXit trial prolonged abstinence: an approximation of the density of the posterior distribution of the odds ratio comparing intervention versus control. (b) NEXit trial point prevalence: an approximation of the density of the posterior distribution of the odds ratio comparing intervention versus control.

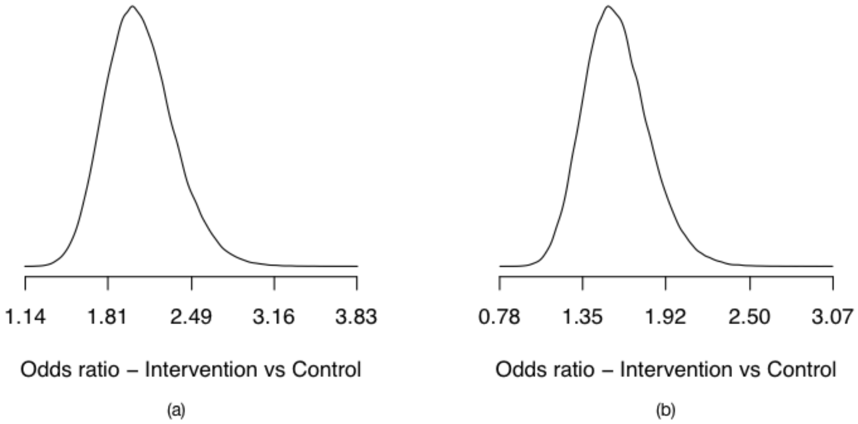


Table 2. Posterior probability of odds ratios at certain thresholds.

Outcome	Odds ratio			
	>1.0	>1.5	>2.0	>2.5
Prolonged abstinence	>99.99%	99.05%	57.37%	7.05%
Point prevalence	99.96%	62.50%	4.19%	0.054%

Comparing Null Hypothesis Significance Testing Versus the Bayesian Approach

It is actually not very easy or straightforward to compare the quantities that the NHST and the Bayesian approach produce. The numbers in Table 1 are in terms of the likelihood of the data—that is, whether the data are extreme given a hypothesized world. The numbers in Table 2 tell us the relative probability among the different worlds directly, given the data that we have collected. While the two approaches may seem to come to the same conclusion in this case—they both agree that the intervention has an effect—it is important to note that the NHST approach only says that the population effect is not 0 and has based this judgment on an arbitrarily chosen threshold, while at the same time imagining that the experiment could be repeated many times. The Bayesian approach says nothing about statistical significance, but rather communicates what we know about the NEXit intervention given the data at hand; it is the researcher's job to transfer the statistical analysis to the real world. It is also the researcher's job to judge the data in light of the model that was chosen, the way the data were collected, existing scientific knowledge, and the novelty of the result. Such things are not meant to be answered by statistical models.

AMADEUS Trial

Much like the NEXit trial, the AMADEUS trial invited college and university students in Sweden to partake in the evaluation of a novel text-based alcohol intervention. The goal was to show that the intervention would reduce alcohol consumption in the group that was given access to the intervention as compared with the control group, who were referred to a website on which they could answer questions about their alcohol consumption and get feedback. The trial ran during the spring of 2016 and included participants who had at least two heavy episodic drinking occasions per month, defined as drinking more than 4 (women) or 5 (men) standard drinks on 1 occasion. The primary outcome measure was the total number of standard drinks consumed per week.

Statistical Model

The outcome measure in the AMADEUS trial was not a coin flip, as there are more than two possible outcomes when asking an individual how many standard drinks they consume per week. Rather, the outcome is a count variable: a variable that can take on values of 0, 1, 2, and so on (participants were not allowed to answer in partial standard drinks). To model this type of data, the researchers decided to use a negative binomial regression model. Just like the logistic regression model used for NEXit has an important quantity known as the odds ratio, the negative binomial regression has a quantity known as the *incident rate ratio* (IRR). This quantity should be interpreted as follows: take the number of standard drinks that the control group drinks on average and multiply by the IRR to get the number of standard drinks that the intervention group drinks on average. Therefore, an IRR of 1 would mean that there was no difference between the groups, less than 1 would mean that the intervention group drank less, and greater than 1 would mean that the intervention group drank more than the control group.

Null Hypothesis Significance Testing of the AMADEUS Trial

From the 896 randomly allocated participants, 816 responses to the primary outcome measure were collected. The IRR was determined using negative binomial regression, and a predefined threshold of .05 was used to determine statistical significance. Table 3 presents the maximum likelihood estimate of IRR, 95% confidence interval, and *P* value. The null hypothesis that the 2 groups consumed the same amount of alcohol after the intervention could not be rejected, since the IRR could not be shown to be significantly different from 1. The population value falls within 95% of all the 95% confidence intervals that can be computed.

Bayesian Analysis of the AMADEUS Trial

As we know by now, the Bayesian approach begins by assigning prior distributions to unknown quantities, and we used flat priors as before (assigning equal probability to all values of the unknown quantities before taking into account the data). Using Bayes' theorem, we computed the posterior distribution over the IRR, depicted in Figure 8 (comparing the total weekly consumption of the intervention group with that of the control group). This is the outcome of the Bayesian analysis, and we can now use this posterior distribution to answer a series of scientific questions.

The AMADEUS trial tested a novel text-based intervention delivered to mobile phones versus referral to a website with a questionnaire and feedback. Let us assume that it was decided that there are certain levels of effect that have real-world implications. For instance, we may define a major preference for the novel intervention if the IRR is less than 0.9 (IRR<0.9), a minor preference if the IRR is between 0.9 and 1.0 (0.9<IRR<1.0), a minor preference for referring to the questionnaire and feedback if the IRR is between 1.0 and 1.1 (1.0<IRR<1.1), and a major preference if the IRR is above 1.1 (IRR>1.1). Table 4 presents these cases as questions, along with their answers. There is a small probability of a major preference for the novel intervention (the posterior probability that IRR was <0.9 is 3.3%). There is a 55.4% probability of a minor preference for the novel intervention, and a 39.6% probability of a minor preference for referring to the questionnaire. Finally, there is a 1.8% probability that referring to the questionnaire should be majorly preferred.

The routine practice at colleges and universities in Sweden is to email all students each year and refer them to the questionnaire and feedback that the control group was offered in the AMADEUS study. Should the novel intervention under trial be considered helpful and replace the questionnaire? It is interesting to note that the original publication [22] discussed potential issues with the study being underpowered and nonsignificance of the hypothesis tests, while the Bayesian approach that we have taken here allows us to discuss the real-world ramifications of the data collected. Based on the data that were collected during the trial and the model that we have chosen, we can say that it is more likely than not that the intervention had a more positive effect in the trial than referral to the questionnaire, but the difference in probability is small. We must therefore assess other factors, including an investigation into the uptake of the two different approaches: an intervention with a small effect that is used by many could be preferred to an intervention with great effect used by few. There are more information technology costs involved in the novel intervention, yet there is less administration from the student health care centers. This type of reasoning must be guided by researchers and experts, and made available to potential users and practitioners so that they take this into account before deciding whether the novel intervention is suitable.

Table 3. Original null hypotheses significance testing of the AMADEUS trial. Incident rate ratio (IRR) is given comparing intervention with control, as per negative binomial regression.

Outcome	IRR	95% CI	*P* value[a]
Weekly alcohol consumption	0.99	0.90-1.09	.83

[a] 2-tailed.

Figure 8. Approximation of the density of the posterior distribution of the incident rate ratio, comparing intervention versus control in the AMADEUS trial.

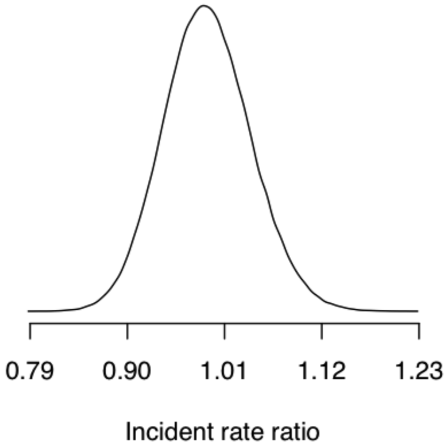


Table 4. Posterior probability of incident rate ratio (IRR) for predefined effect levels.

Outcome	IRR<0.9	0.9<IRR<1.0	1.0<IRR<1.1	IRR>1.1
Weekly alcohol consumption	3.3%	55.4%	39.6%	1.8%

Comparing Null Hypothesis Significance Testing Versus the Bayesian Approach

The NHST analysis presented in Table 3 does not reject the null hypothesis; thus, the analysis cannot say anything about the effect of the novel intervention. Crucially, it does not tell us what the probability of the intervention having a positive or negative effect is, but rather the maximum likelihood estimate is just a sample from a sampling distribution for which we do not know the mean (the population value). It is unfortunate that, due to a conventionally decided threshold of .05 and a test against a very strict null hypothesis, the analysis leads us to a dead end from where we cannot express much more about the intervention.

The Bayesian analysis gives us a posterior distribution, and then the scientific inference can begin. Scientific inference cannot rely on conventional thresholds applied across all research fields, but rather scientific inference must be based on the real-world context and study parameters. The levels we choose to assess the effect can be understood by readers because these chosen levels have direct real-world implications—no such connection can be made to a .05 *P* value threshold.

Discussion

Null Hypothesis Significance Testing and Bayesian Analysis

Setting aside the mathematical differences between the two approaches, the most prominent difference is perhaps that the Bayesian approach put forward here does not incorporate the same type of null hypothesis testing that is so strongly rooted in conventional practice. This ties into the fact that the output from the Bayesian analysis is the posterior distribution over the parameters of our model. Therefore, the Bayesian approach does not attempt to identify a fixed value for the parameters and dichotomize the world into significant and nonsignificant, but rather relies on the researcher to do the scientific inference and not to delegate this obligation to the statistical model. It should not be forgotten that all statistical inference is based on a model, whether we take the NHST or Bayesian approach, and that these models are approximations of the real world. In both cases, there needs to be a leap of faith that the model chosen is a good enough approximation. We should therefore be careful not to let the model alone make assessments of the bias of the coin, but rather we must take what the model tells us and then go back to the real world and do the scientific inference ourselves.

We expect researchers to add their interpretation of their results, grounded in previous studies and current theory, balanced with cost and benefit, etc. We have purposely kept short the analyses that we have presented, but a full analysis cannot end with a posterior distribution; some scientific inference needs to be conducted. One attractive aspect of the Bayesian analyses that we have conducted herein is the way in which we ask questions of the models that have been created. For instance, the questions in Table 2 relate to the odds ratio, a quantity that can be interpreted with a real-world meaning. Compare this with an arbitrary threshold for the *P* value, which only applies in the null hypothesis world, and even then is difficult to connect to a real-world quantity.

In the NHST approach, we are assessing the population value, and we state upfront our intentions: if the null hypothesis is rejected, then we will say that the coin is biased. In this sense, we are giving a license to the statistical model to do scientific inference. Once the analysis is complete and the null hypothesis is rejected, we are not much wiser about the population value; as we have discussed, confidence intervals are not as good an indication of the location of the fixed population value as we might think. In case the hypothesis is not rejected; we have very little use of our analysis. Furthermore, the NHST approach is rooted in the idea of being able to redo the experiment many times (so as to get a sampling distribution). Even if we can rely on theoretical results to get this sampling distribution without actually going back in time and redoing the experiment, the underlying idea can be somewhat problematic. What do we mean by redoing an experiment? Can we redo a randomized controlled trial while keeping all things equal and recruiting a new sample from the study population? We might just overlook this philosophical obstacle if we like, but we should not forget that we are asking our statistical models to use such an assumption to make dichotomous decisions.

The Bayesian analysis outputs a posterior distribution, which then must be used to assess whether the coin is fair. We can say something about the value of the quantity of interest given our data, since the posterior distribution is a distribution over all possible values of the quantity. There exist Bayesian hypothesis frameworks that allow for a systematic way of making dichotomous decisions, and the interested reader may want to look into the field of decision theory, but at the end of the day the researcher must use the posterior distribution to assess the real-world implications. Imagine that we were assessing whether a medical procedure would be beneficial for a patient. We would have to weigh this probability with the risk for the patient: a 95% probability in favor of the procedure may be necessary if the procedure is invasive (eg, surgery), while a 60% probability in favor of the procedure may be okay if it simply involves a patient taking part in a seminar.

Prior Distributions

It is usually the prior distribution that is contested by non-Bayesian proponents. How can we know anything about a parameter before we collect any data? While it is not made explicit, the non-Bayesian approach does in a sense assume flat priors on all parameters, which is why many newcomers to the Bayesian field feel that flat priors should be used all the time. However, the belief that flat priors are objective because they assign the same probability to all outcomes is not well grounded. Consider, for instance, the NEXit trial, where we used flat priors, which encodes that before we analyze the data we believe that all outcomes are equally likely. This is, however, subjective: believing it equally likely that 20% to 25% in the intervention group will quit smoking and that 90% to 95% will quit smoking. We know that brief interventions usually have a small to moderate effect size; thus, assuming a flat prior is a subjective choice going against what is known. Therefore, subjective modelling choices are unavoidable, regardless of whether one takes the Bayesian approach. The fact that the Bayesian approach requires researchers to explicitly state their prior beliefs is actually a boon, since it forces us to be explicit about this choice, rather than hiding it. Had this paper focused solely on the analysis of the NEXit and AMADEUS trials, we would have followed the suggestion of Spiegelhalter et al [23] to conduct our analysis under several priors, one that encodes indifference, one that encodes the genuine opinion among practitioners, and one that encodes skepticism toward the new intervention. It should also be noted, as McShane et al [3] pointed out, that while using a *P* value threshold may seem like a way to break subjective interpretations of statistical analyses, *P* values are highly subjective in the sense that the choice of which models to use, which covariates to include, which tests to perform, etc, all produce different *P* values.

Interpreting Results

Practitioners, patients, the media, journal editors, and reviewers are keen to ask “does it work?” or “is it significant?” It is of course convenient to tell a patient that an intervention has been proven to have effect in a scientific study, but such statements are vague at best and lying at worst, and are still based on statistical models with arbitrarily decided-upon thresholds and null hypotheses. We should be communicating the probability that the intervention effect lies within a given range, such as that the odds ratio is greater than 1. Practitioners, patients, the media, journals, and reviewers can then use their own situation and expertise to assess the implications. We can take the posterior distribution and set it into economic and social contexts. An intervention with a 75% probability of a positive effect may still be defensible to implement, since it may be very cheap and noninvasive, while an intervention that has 95% probability of a positive effect might not be economically feasible to implement. Once we remove ourselves from the dichotomization of evidence, other things start to take precedence: critically assessing the models chosen, evaluating the quality of the data, interpreting the real-world impact of the results, etc.

We argue that the dichotomization, or be it trichotomization, is more misleading and misunderstood than Amrhein and Greenland [2] and McShane et al [3] pointed out. Many researchers and readers of scientific literature interpret statistical significance as true and nonstatistical significance as false, but this dichotomization does not exist, since statistical significance splits the world into a true state within which there exists an effect and a state in which there is ambivalence, which is not the opposite of the true state. It is not a not-true state and not a false state. Thinking in terms of statistical significance leads to a very difficult to understand dichotomization. The proposal from Benjamin et al [1] would further complicate matters, as we would end up in an even more difficult to understand trichotomization, and it raises the question of whether scientific discoveries based on *P* value thresholds of .05 from the past should now be considered nonsignificant.

Conclusion and Call for Papers

While, compared with the NHST approach, the use of Bayesian methods to analyze randomized controlled trials is virtually nonexistent, it has increased over the past few years (Lee and Chu [24]). As further evidence of the traction Bayesian methods are achieving, the US Food and Drug Administration has released guidelines for the use of Bayesian statistics in medical device clinical trials [25].

It may yet be some time until all trials report Bayesian posteriors with scientific inference; it is nevertheless time to both educate researches about Bayesian methods and include these methods alongside current practice. The *Journal of Medical Internet Research* has issued a call for papers for a special theme issue that will be dedicated to the (re-)analysis of data from randomized controlled trials using a Bayesian framework. We invite researchers to reanalyze data from their previously published trials and write a short paper about their new analysis. Please see the call for papers on JMIR’s website (https://www.jmir.org/announcement/view/172) for further details.

Conflicts of Interest

MB owns a private company that develops and distributes evidence-based lifestyle interventions to be used in health care settings.

References

1. Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers E, Berk R, et al. Redefine statistical significance. Nat Hum Behav 2017 Sep 1;2(1):6-10. [doi: 10.1038/s41562-017-0189-z]
2. Amrhein V, Greenland S. Remove, rather than redefine, statistical significance. Nat Hum Behav 2017 Sep 25;2(1):4-4. [doi: 10.1038/s41562-017-0224-0]
3. McShane BB, Gal D, Gelman A, Robert C, Tackett JL. Abandon statistical significance. arXiv:1709.07588. 2017 Sep 21. URL: https://arxiv.org/pdf/1709.07588v1.pdf [WebCite Cache ID 72VU8r4KP]
4. Wasserstein RL, Lazar NA. The ASA's statement on p-values: context, process, and purpose. Am Stat 2016 Jun 09;70(2):129-133. [doi: 10.1080/00031305.2016.1154108]
5. Trafimow D, Marks M. Editorial. Basic Appl Soc Psychol 2015 Feb 12;37(1):1-2. [doi: 10.1080/01973533.2015.1012991]
6. Nuzzo R. Scientific method: statistical errors. Nature 2014 Feb 13;506(7487):150-152. [doi: 10.1038/506150a] [Medline: 24522584]
7. McShane BB, Gal D. Statistical significance and the dichotomization of evidence. J Am Stat Assoc 2017 Oct 30;112(519):885-895. [doi: 10.1080/01621459.2017.1289846]
8. Berry D. A P-value to die for. J Am Stat Assoc 2017 Oct 30;112(519):895-897. [doi: 10.1080/01621459.2017.1316279]

9. Briggs WM. The substitute for P-values. J Am Stat Assoc 2017 Oct 30;112(519):897-898. [doi: 10.1080/01621459.2017.1311264]
10. Gelman A, Carlin J. Some natural solutions to the P-value communication problem–and why they won't work. J Am Stat Assoc 2017 Oct 30;112(519):899-901. [doi: 10.1080/01621459.2017.1311263]
11. Laber EB, Shedden K. Statistical significance and the dichotomization of evidence: the relevance of the ASA Statement on Statistical Significance and p-values for statisticians. J Am Stat Assoc 2017 Oct;112(519):902-904 [FREE Full text] [doi: 10.1080/01621459.2017.1311265] [Medline: 29348701]
12. McShane BB, Gal D. Rejoinder: statistical significance and the dichotomization of evidence. J Am Stat Assoc 2017 Oct 30;112(519):904-908. [doi: 10.1080/01621459.2017.1323642]
13. Browne EN, Rathinam SR, Kanakath A, Thundikandy R, Babu M, Lietman TM, et al. A Bayesian analysis of a randomized clinical trial comparing antimetabolite therapies for non-infectious uveitis. Ophthalmic Epidemiol 2017 Dec;24(1):63-70 [FREE Full text] [doi: 10.1080/09286586.2016.1255764] [Medline: 27982726]
14. Goodman SN, Sladky JT. A Bayesian approach to randomized controlled trials in children utilizing information from adults: the case of Guillain-Barré syndrome. Clin Trials 2005;2(4):305-10; discussion 364. [doi: 10.1191/1740774505cn102oa] [Medline: 16281429]
15. Kruschke JK. Bayesian estimation supersedes the t test. J Exp Psychol Gen 2013 May;142(2):573-603. [doi: 10.1037/a0029146] [Medline: 22774788]
16. Morris RK, Malin GL, Quinlan-Jones E, Middleton LJ, Diwakar L, Hemming K, et al. The Percutaneous shunting in Lower Urinary Tract Obstruction (PLUTO) study and randomised controlled trialvaluation of the effectiveness, cost-effectiveness and acceptability of percutaneous vesicoamniotic shunting for lower urinary tract obstruction. Chapter 7, Bayesian analysis of the randomised controlled trial. Health Technology Assessment, No. 17.59. Southampton, UK: NIHR Journals Library; 2013 Dec. URL: https://www.ncbi.nlm.nih.gov/books/NBK261381/ [WebCite Cache ID 72VUYhn1I]
17. Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian Approaches to Trials and Health-Care Evaluation. New York, NY: John Wiley & Sons, Ltd; 2003.
18. Wijeysundera DN, Austin PC, Hux JE, Beattie WS, Laupacis A. Bayesian statistical inference enhances the interpretation of contemporary randomized controlled trials. J Clin Epidemiol 2009 Jan;62(1):13-21.e5. [doi: 10.1016/j.jclinepi.2008.07.006] [Medline: 18947971]
19. Carvalho CM, Polson NG, Scott JG. Handling sparsity via the horseshoe. 2009 Presented at: 12th International Conference on Artificial Intelligence and Statistics (AISTATS); Apr 16-18, 2009; Clearwater Beach, FL, USA.
20. Piironen J, Vehtari A. Sparsity information and regularization in the horseshoe and other shrinkage priors. Electron J Stat 2017;11(2):5018-5051. [doi: 10.1214/17-EJS1337SI]
21. Müssener U, Bendtsen M, Karlsson N, White IR, McCambridge J, Bendtsen P. Effectiveness of short message service text-based smoking cessation intervention among university students: a randomized clinical trial. JAMA Intern Med 2016 Mar;176(3):321-328. [doi: 10.1001/jamainternmed.2015.8260] [Medline: 26903176]
22. Thomas K, Müssener U, Linderoth C, Karlsson N, Bendtsen P, Bendtsen M. Effectiveness of a text messaging-based intervention targeting alcohol consumption among university students: a randomized controlled trial. JMIR Mhealth Uhealth 2018 Jun 25;6(6):e146 [FREE Full text] [doi: 10.2196/mhealth.9642] [Medline: 29941417]
23. Spiegelhalter DJ, Freedman LS, Parmar MKB. Applying Bayesian ideas in drug development and clinical trials. Stat Med 1993 Aug;12(15-16):1501-1511. [doi: 10.1002/sim.4780121516]
24. Lee JJ, Chu CT. Bayesian clinical trials in action. Stat Med 2012 Nov 10;31(25):2955-2972 [FREE Full text] [doi: 10.1002/sim.5404] [Medline: 22711340]
25. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Devices and Radiological Health, Division of Biostatistics, Office of Surveillance and Biometrics. Guidance for the use of Bayesian statistics in medical device clinical trials. 2010 Feb 5. URL: https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071121.pdf [WebCite Cache ID 72VVNIrt8]

Abbreviations

IRR: incident rate ratio
NHST: null hypothesis significance testing

Edited by G Eysenbach; submitted 26.04.18; peer-reviewed by J Luo, Z Dienes; comments to author 03.08.18; revised version received 07.08.18; accepted 07.08.18; published 24.10.18.

Please cite as:
Bendtsen M
A Gentle Introduction to the Comparison Between Null Hypothesis Testing and Bayesian Analysis: Reanalysis of Two Randomized Controlled Trials
J Med Internet Res 2018;20(10):e10873
URL: http://www.jmir.org/2018/10/e10873/
doi:10.2196/10873
PMID:30148453

Original Paper

Predicting Adherence to Internet-Delivered Psychotherapy for Symptoms of Depression and Anxiety After Myocardial Infarction: Machine Learning Insights From the U-CARE Heart Randomized Controlled Trial

John Wallert[1], MSc; Emelie Gustafson[1], MA; Claes Held[2,3], MD, PhD; Guy Madison[4], PhD; Fredrika Norlund[1], PhD; Louise von Essen[1], PhD; Erik Martin Gustaf Olsson[1], PhD

[1]Clinical Psychology in Healthcare, Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden

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

[3]Department of Medical Sciences, Uppsala University, Uppsala, Sweden

[4]Department of Psychology, Umeå University, Umeå, Sweden

Corresponding Author:
John Wallert, MSc
Clinical Psychology in Healthcare
Department of Women's and Children's Health
Uppsala University
Dag Hammarskjölds väg 14B
SE-75185
Uppsala,
Sweden
Phone: 46 729999217
Email: john.wallert@kbh.uu.se

Abstract

Background: Low adherence to recommended treatments is a multifactorial problem for patients in rehabilitation after myocardial infarction (MI). In a nationwide trial of internet-delivered cognitive behavior therapy (iCBT) for the high-risk subgroup of patients with MI also reporting symptoms of anxiety, depression, or both (MI-ANXDEP), adherence was low. Since low adherence to psychotherapy leads to a waste of therapeutic resources and risky treatment abortion in MI-ANXDEP patients, identifying early predictors for adherence is potentially valuable for effective targeted care.

Objectives: The goal of the research was to use supervised machine learning to investigate both established and novel predictors for iCBT adherence in MI-ANXDEP patients.

Methods: Data were from 90 MI-ANXDEP patients recruited from 25 hospitals in Sweden and randomized to treatment in the iCBT trial Uppsala University Psychosocial Care Programme (U-CARE) Heart study. Time point of prediction was at completion of the first homework assignment. Adherence was defined as having completed more than 2 homework assignments within the 14-week treatment period. A supervised machine learning procedure was applied to identify the most potent predictors for adherence available at the first treatment session from a range of demographic, clinical, psychometric, and linguistic predictors. The internal binary classifier was a random forest model within a 3×10–fold cross-validated recursive feature elimination (RFE) resampling which selected the final predictor subset that best differentiated adherers versus nonadherers.

Results: Patient mean age was 58.4 years (SD 9.4), 62% (56/90) were men, and 48% (43/90) were adherent. Out of the 34 potential predictors for adherence, RFE selected an optimal subset of 56% (19/34; Accuracy 0.64, 95% CI 0.61-0.68, P<.001). The strongest predictors for adherence were, in order of importance, (1) self-assessed cardiac-related fear, (2) sex, and (3) the number of words the patient used to answer the first homework assignment.

Conclusions: For developing and testing effective iCBT interventions, investigating factors that predict adherence is important. Adherence to iCBT for MI-ANXDEP patients in the U-CARE Heart trial was best predicted by cardiac-related fear and sex, consistent with previous research, but also by novel linguistic predictors from written patient behavior which conceivably indicate verbal ability or therapeutic alliance. Future research should investigate potential causal mechanisms and seek to determine what underlying constructs the linguistic predictors tap into. Whether these findings replicate for other interventions outside of Sweden, in larger samples, and for patients with other conditions who are offered iCBT should also be investigated.

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(10):e10754) doi:10.2196/10754

KEYWORDS

applied predictive modeling; cardiac rehabilitation; linguistics; supervised machine learning; recursive feature elimination; treatment adherence and compliance; Web-based interventions

Introduction

Myocardial infarction (MI) afflicts more than 7 million individuals each year, making it the most common acute cardiac event caused by cardiovascular disease (CVD)—the leading cause of death in the world [1]. After an acute MI, behavior changes are required in order to reduce the risk of reinfarction, stroke, and death. Important health-promoting behaviors include smoking cessation, regular physical activity, a healthy diet, and adherence to medications [2,3].

A substantial subgroup of patients with MI additionally also suffer from symptoms of anxiety, depression, or both (MI-ANXDEP). MI-ANXDEP patients have a higher risk factor burden and worse prognosis compared to MI patients in general [4,5]. Alongside prescribed physical activity, psychological support is therefore suggested as treatment for MI-ANXDEP patients to reduce affective symptoms [6-8] and in turn facilitate health-promoting behavioral change toward cardiac risk reduction [2,9]. Psychological support in the form of cognitive behavior therapy (CBT) has shown effectiveness on psychological symptoms for several common psychiatric disorders. Internet-delivered CBT (iCBT) is a cost-effective version of face-to-face CBT [10,11] that, however, places high demands on the reading and writing abilities of the patient. Patient dropout from iCBT in a meta-analysis for depression (n=40 studies) was 57%. Subanalyses showed 28% and 38% dropout from therapist- and administrator-supported iCBT, respectively. These attrition numbers are substantial, suggesting further research into adherence to iCBT. Although adherence to iCBT is not a guarantor for iCBT effectiveness, adherence is arguably a prerequisite for possible iCBT effect and thus worthwhile to study in its own right [12].

The multicenter Uppsala University Psychosocial Care Programme (U-CARE) Heart study was the first randomized controlled trial to test the effectiveness of a therapist-supported iCBT treatment for MI-ANXDEP patients [13,14]. The U-CARE Heart trial design arguably had high ecological (clinical routine) validity [14] compared to other iCBT trials that have relied on self-referral and applied stricter inclusion/exclusion criteria [15,16]. U-CARE Heart also featured relatively low adherence to iCBT, which in turn lacked effect at the group comparison level [13]. For future dissemination of iCBT, it is crucial to assess the effect and practical utility of iCBT under ecologically valid conditions [17] such as in U-CARE Heart and explore factors that predict adherence if low adherence is a problem in such contexts. Adherence to treatment by cardiovascular patients has been thoroughly investigated with respect to medical compliance [18] but not with respect to iCBT offered to MI-ANXDEP patients.

Treatment adherence is in general a multifactorial phenomenon. Adherence to and effectiveness of iCBT has been associated with higher education, older age, and female sex [19,20]. In addition to these background predictors, both patient motivation [12] and treatment credibility [19] have been found to substantially augment adherence to iCBT. Regarding MI-ANXDEP psychological symptomatology, patient motivation and belief in the iCBT treatment are probably also predicated on cardiac-related anxious and depressive symptom severity as well as placebo priors regarding iCBT effectiveness. The so-called therapeutic alliance, the patient-therapist bond sought to be developed during individual psychotherapy, has also been found to benefit adherence to iCBT [12]. Furthermore, it is worthwhile to investigate the relative predictive power of some cardiovascular variables, as somatic disease severity might also influence adherence to iCBT among MI-ANXDEP patients.

The present iCBT U-CARE Heart study design offered a group of additional predictors that have not been assessed in this way, namely linguistic variables based on the texts that patients wrote in response to their standardized homework assignments. Syntactic structure and word use has to some extent been investigated before with regard to anxiety and depression [21-23], and number of words used when applying for Web-based depression treatment has been shown to correlate with adherence [24]. In the U-CARE Heart study, the texts are logged at the start of treatment, and various quantitative variables can be extracted from these texts using linguistic procedures. These extracts were then modeled as additional linguistic predictors for adherence in our study. It is likely that more verbally oriented and engaged patients write longer and more complex texts and also adhere better to verbally demanding treatments such as iCBT. It is also possible that these linguistic predictors to some extent are proxies of other established predictors for adherence (eg, motivation, treatment credibility, and therapeutic alliance) and as proxies would thus hold predictive power. We propose that these linguistic predictors might contribute to the acuity of predictive models in addition to known predictors of iCBT adherence (eg, education, age, sex, and symptom severity).

The objective of our study was to investigate if predictors available up to the start of treatment (initial homework assignment response) would predict adherence to iCBT treatment at first follow-up in MI-ANXDEP patients. To this end, we applied a contemporary machine learning procedure to

U-CARE Heart data to manage the relatively large amount of predictors and complex covariance structure. We hypothesized that symptom severity, age, sex, education, and linguistic behavior would predict adherence to treatment. We also hypothesized that more severe symptoms, younger age, being a woman, having a higher education, and using more words in the assignment response would be positively associated with adherence to iCBT.

Methods

Treatment and Study Sample

The recruitment, treatment, and follow-up of patients has been described in detail elsewhere [13,14,25]. In summary, the trial recruited 239 patients from 25 Swedish hospitals and randomized 122 patients to a control group and 117 patients to therapist-guided and self-tailored 14-weeks of iCBT. Of these 117 patients, 27 did not respond to any homework assignments and were excluded due to lack of data on all linguistic variables. This rendered a study sample of 90 patients. The treatment modules consist of homework assignments to be completed by the patient on which the licensed psychologist provided feedback. The psychologist communicated with the patient through an in-portal message system. The first two homework assignments were standardized for all patients. This standardization removed the problem of complex patient-psychologist interactions that are inherently dynamic. After the first two assignments, the treatment was self-tailored. The treatment consisted of psychoeducation on principals for rational versus irrational thinking, graded exposure to fearful stimuli, the negative feedback loop in depressive behavior, as well as relaxation training, improving communication skills, additional behavioral change toward long-term goals, and relapse prevention.

Outcome and Initial Predictor Selection

The outcome variable was dichotomous: adherence was defined as completing 3 or more homework assignments (≥21% of total treatment), and nonadherence was defined as having completed less than that. This cutoff was chosen in part because it is clinically relevant to ascertain who continues with the self-tailored part of the U-CARE Heart treatment after completing the initial 2 standardized homework assignments versus who does not continue. Furthermore, the chosen cutoff rendered fairly balanced classes for the machine learning procedure, which is important for it to work properly with moderately sized data [26]. Psychological (EO, JW, FN), cardiologic (CH), and linguistic (EG) experts selected an initial set of 34 possible predictors of psychometric, linguistic, clinical, and demographic type. See Table 1 for further details on the predictors.

Linguistic Predictors

The linguistic predictors were extracted from the patients' answers to the first standardized homework assignment, which consisted of an introductory text and 8 questions designed for the patient to describe their MI, associated psychological reaction, present psychological state, present social support, and what the patient wanted from iCBT treatment. In effect, patients had access to the same material prior to carrying out their homework assignment [13,14]. Since the patients had read both example answers and an introductory text before writing their response, it is possible that the patients' choice of words would be substantially, but also equally, primed when answering the questions. The linguistic factors investigated were (1) the number of words used, (2) average sentence length, (3) normalized frequencies (results given as n/1000 words) of adjectives or adverbs, (4) normalized frequencies of possessive pronouns, (5) normalized frequencies of personal pronouns, (6) whether or not the patient mentions the MI, and the (7) frequency of mutual usage of a small set of prespecified key words (used both in a standardized question and in a patient answer). Predictors 1 through 7 were selected on the basis of them being possibly indicative of adherence to iCBT as probable proxies for verbal skill, socioeconomic status, and investment in therapy, all arguably important factors for iCBT adherence. See Multimedia Appendix 1 for further details on the linguistic predictors.

Imputation

Five of the 34 predictors had missing data, in the order of proportion missing: number of standard glasses of alcohol consumed per week, 11% (10/90); BMI, 10% (9/90); heart rate, 7% (6/90); systolic blood pressure (SBP), 7% (6/90); and the number of days between hospital admission for MI and study randomization, 4% (4/90). Missing values were thus relatively few and not considered missing completely at random (MCAR), instead their missingness was assumedly related to the other measured variables (MAR). We also did not impute the outcome. Thus, k nearest neighbor (k-NN) imputation was performed with number of nearest cases (k) set to 3 and all variables with missing values imputing the median of k values. The k-NN is a well-established algorithm for imputing both numerical and categorical variables based on a generalized distance metric [33,34]. In this study, the Hower distance metric was used [35]. If k, from which the algorithm borrows values for cases with missing values, is set low (eg, $k \leq 3$), imputation with k-NN also preserves much of the underlying correlational structure of data.

Predictive Modeling

Adherence is a multifactorial problem [18,20], which suggested a multivariable prediction model. For testing the relative power of predictors, a useful method would be one that can weigh the variables according to their relative importance for solving the binary classification problem of predicting adherence versus nonadherence. The Breiman random forest model [36,37] is a well-established ensemble method which usually performs well with moderately sized data, is insensitive to multicollinearity and nuisance variables, and has previously worked well with MI patient data [38]. These model characteristics are suitable for the multiple highly correlated psychometric measures and 90 MI-ANXDEP patients in this study. Random forest also models linear and higher-order effects automatically, which concurs with the main study objective to estimate the total relative importance of a range of predictors.

Table 1. Descriptive statistics for all treated patients with myocardial infarction and stratified by adherence to internet-delivered cognitive behavioral therapy.

Variables	All (n=90)	Adherers (n=43)	Nonadherers (n=47)	*P* value	Missing
Demographics					
Age (years) mean (SD)	58.4 (9.4)	57.0 (10.4)	60.0 (8.3)	.17	0
Women, n (%)	34 (38)	23 (54)	11 (23)	.006	0
Civic status, n (%)				.80	0
Single	15 (17)	8 (19)	7 (15)		
Cohabitant/married	72 (80)	34 (79)	38 (81)		
Not single but living alone	3 (3)	1 (2)	2 (4)		
Education (highest attained) n (%)				.79	0
Elementary	14 (16)	5 (12)	9 (19)		
High school	31 (34)	16 (37)	15 (32)		
University ≤3 years	20 (22)	10 (23)	10 (21)		
University >3 years	25 (28)	12 (28)	13 (28)		
Country of birth, n (%)	17 (19)	8 (19)	9 (19)	>.99	0
Clinical					
Heart rate, mean (SD)	77.0 (20.4)	77.6 (21.3)	76.5 (19.7)	.81	6
SBP[a], mean (SD)	149.5 (32.0)	150.5 (28.2)	148.5 (35.6)	.78	6
BMI[b], mean (SD)	27.9 (5.0)	27.9 (5.8)	28.0 (4.3)	.89	9
Alcohol (glasses/week), median (IQR[c])	2.0 (0.0, 7.3)	2.0 (0.0, 8.5)	2.0 (0.0, 5.0)	.44	10
Current smoker, n (%)	4 (4)	2 (5)	2 (4)	>.99	0
CVD[d] medication adherence, n (%)	18 (20)	11 (26)	7 (15)	.32	0
Psychoactive medication, n (%)				.45	0
None	75 (83)	34 (79)	41 (88)		
As needed	6 (7)	3 (7)	3 (6)		
Regularly	7 (8)	4 (9)	3 (6)		
Regularly and as needed	2 (2)	2 (5)	0 (0)		
Other current counseling, n (%)				.48	0
No	67 (74)	31 (72)	36 (77)		
≥Once per year, <once per month	9 (10)	6 (14)	3 (6)		
≥Once per month	14 (16)	6 (14)	8 (17)		
Psychometric, mean (SD)					
CAQ[e] fear	12.7 (6.0)	14.6 (5.4)	11.0 (6.0)	.004	0
CAQ avoidance	7.3 (4.4)	7.4 (4.2)	7.1 (4.7)	.74	0
CAQ attention	5.7 (3.2)	6.4 (3.4)	5.1 (3.0)	.05	0
CAQ total	25.7 (10.0)	28.4 (9.8)	23.2 (9.6)	.01	0
ESSI[f] total	20.1 (4.4)	20.4 (4.0)	19.7 (4.7)	.49	0
EQ5D[g] VAS[h]	66.0 (16.8)	64.7 (15.6)	67.2 (17.9)	.48	0
EQ5D emotional distress	1.0 (0.5)	1.0 (0.5)	1.0 (0.4)	.84	0
MADRS[i] total	14.9 (6.2)	14.9 (5.7)	15.0 (6.7)	.96	0
BADS[j] total	21.4 (6.1)	22.4 (5.7)	20.6 (6.3)	.15	0

Variables	All (n=90)	Adherers (n=43)	Nonadherers (n=47)	*P* value	Missing
HADS[k] anxiety	10.3 (3.0)	10.5 (2.7)	10.2 (3.2)	.71	0
HADS depression	7.9 (3.0)	8.0 (2.7)	7.9 (3.4)	.92	0
HADS total	18.3 (4.7)	18.4 (4.0)	18.2 (5.3)	.77	0
Linguistic					
Number of words, mean (SD)	306.8 (246.7)	376.8 (257.2)	242.7 (220.5)	.009	0
Number of mutual words, mean (SD)	6.2 (5.7)	7.6 (5.9)	4.9 (5.2)	.02	0
Sentence length, mean (SD)	13.0 (5.5)	13.6 (5.0)	12.4 (5.9)	.28	0
Adjectives/adverbs, mean (SD)	193.2 (43.6)	187.4 (39.9)	198.5 (46.6)	.23	0
Possessive pronouns, mean (SD)	13.1 (10.0)	12.8 (8.1)	13.4 (11.5)	.78	0
Personal pronouns, mean (SD)	64.6 (27.1)	70.2 (24.3)	59.4 (28.8)	.06	0
Mentions the MI[l], n (%)	69 (77)	35 (81)	34 (72)	.44	0
Other					
Days from MI to allocation, mean (SD)	70.5 (14.9)	70.3 (15.0)	70.7 (14.9)	.91	4
Way of preferred contact, n (%)				.59	0
Email	63 (70)	29 (67)	34 (72)		
Telephone	11 (12)	5 (12)	6 (13)		
SMS[m]	15 (17)	9 (21)	6 (13)		
Mail	1 (1)	0 (0)	1 (2)		

[a] SBP: systolic blood pressure.

[b] BMI: body mass index.

[c] IQR: interquartile range.

[d] CVD: cardiovascular disease.

[e] CAQ: Cardiac Anxiety Questionniare [27].

[f] ESSI: ENRICHD Social Support Instrument [28].

[g] EQ5D: European Quality of Life Questionnaire–Five Dimensions.

[h] VAS: visual analog scale.

[i] MADRS: Montgomery-Asberg Depression Rating Scale [29,30].

[j] BADS: Behavioral Activation for Depression Scale–Short Form [31].

[k] HADS: Hospital Anxiety and Depression Scale [32].

[l] MI: myocardial infarction.

[m] SMS: short message service.

Although random forest already has built-in cross-validation control for overfitting through its “out-of-bag” predictions, we added a second wrapper layer around the classifier in the form of backwards algorithmic predictor selection via recursive feature elimination (RFE) resampled with 3×10–fold cross-validation [39]. This was done to further decrease the risk of overfitting and remove human bias from the final feature selection. Regular k-fold cross-validation partitions data into k parts and then trains the model k times, each time withholding data belonging to one of the folds and testing each trained model on the corresponding hold-out fold. Modeling results are thereafter usually averaged across resampling folds. Repeated cross-validation is an extension of regular k-fold cross-validation where data is again randomly partitioned into k-folds for each pass of regular cross-validation. Since random forest was used as the classifier within RFE resampling, the process optimized on classification accuracy, and predictors were ranked on their reduction in node impurity (Gini importance) across decision trees in the random forest ensemble.

Additional Statistics

If not stated differently, we report categorical variables as count (%), numerical variables as arithmetic mean (SD), *P* value for bivariate tests of significance set at 5%, and prediction accuracy for the binary outcome (adherent vs nonadherent) with 95% confidence intervals.

Coding

The linguistic data preprocessing was carried out with the corpus tool AntConc version 3.4.4m (Waseda University) [40], a corpus toolkit for concordancing and text analysis. Linguistic data was also annotated with a *Part of Speech*- tagger for Swedish called Stagger (Stockholm University) [41]. Analysis was done in R version 3.4.0 (The R Foundation for Statistical Computing) [42] using packages *caret, data.table, foreign, ggplot2, ggpubr, ggthemes*, *mice*, *scales*, *tableone*, and *VIM*.

Results

Descriptive data are available in Table 1. Patients who were adherent to iCBT were more frequently women and had higher self-rated cardiac anxiety and cardiac anxiety specifically related to fear and attention compared to those nonadherent. Adherent patients also used more words and more mutual words in their homework assignment. There was a tendency for adherence to increase with age and higher self-rated depression. There were no significant differences between adherers and nonadherers regarding educational attainment, whether Swedish-born or not, civil status, educational attainment, clinical characteristics, days from MI to treatment allocation, or preferred way of contact.

After imputation, the RFE feature selection procedure was applied to extract the most potent predictors for classifying adherers versus nonadherers. Figure 1 shows the resampled result optimized on prediction accuracy and the final optimal model as selected by RFE. This final model used 56% (19/34) of the provided predictors and performed significantly better than did a random model (Accuracy 0.64, 95% CI 0.61-0.68, *P*<.001) although with remaining room for acuity improvement.

Figure 2 plots the main result with each of the 19 top predictors according to RFE by their resampled relative importance for classifying adherers versus nonadherers, showing that the 6 most potent predictors were Cardiac Anxiety Questionnaire (CAQ) fear, sex, number of words, CAQ total, average sentence length, and number of mutual words.

Figure 1. Predictor selection result with recursive feature elimination.

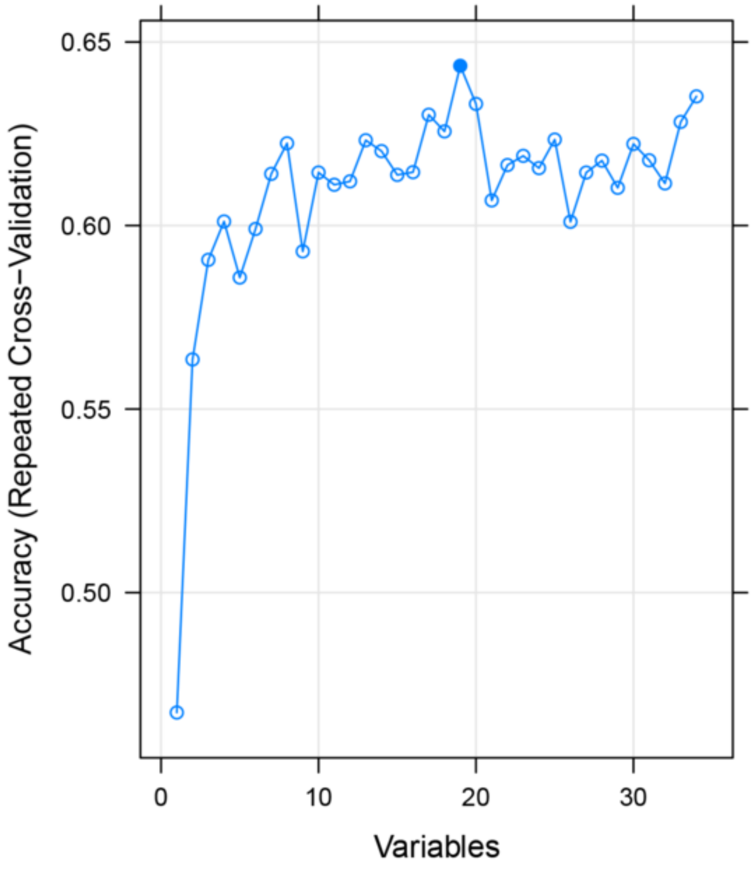
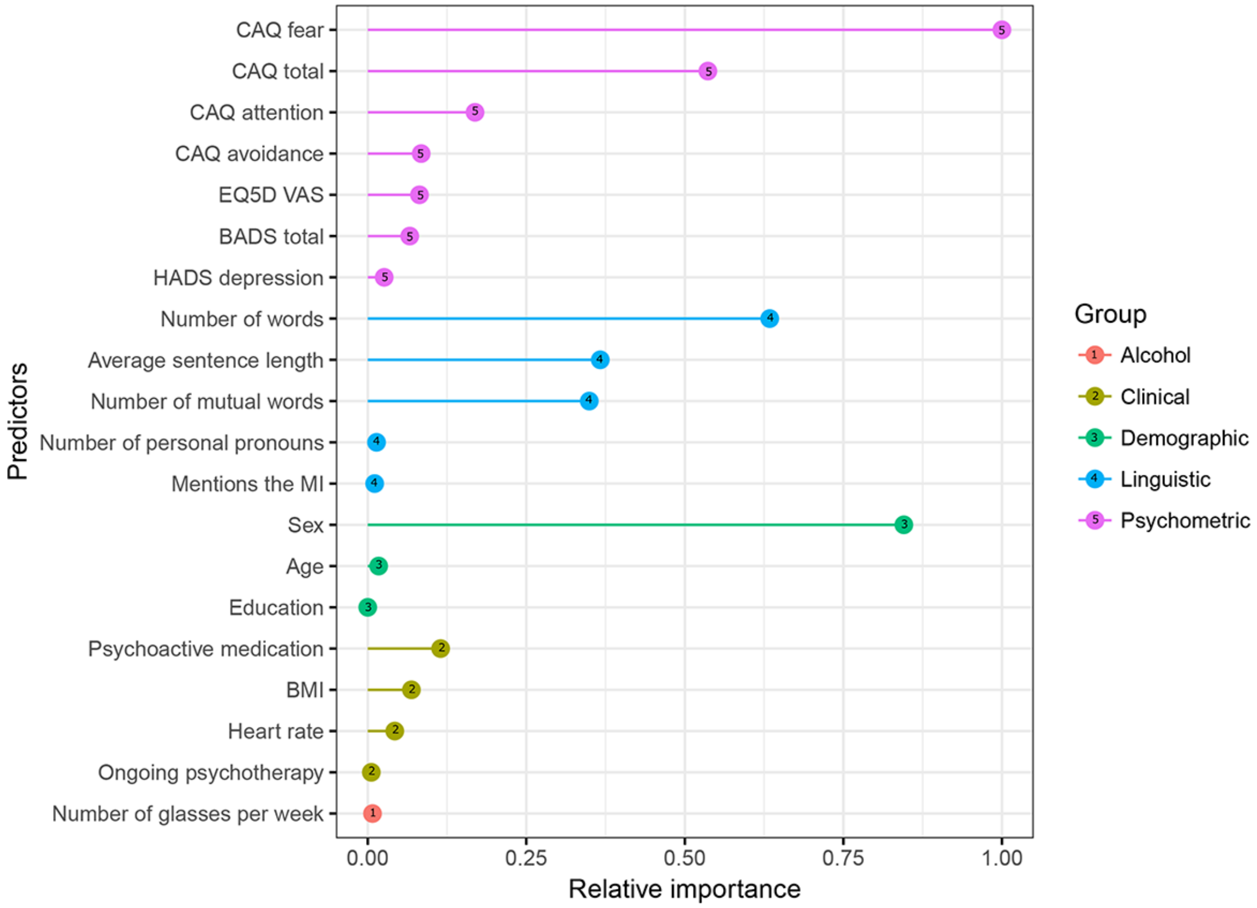


Figure 2. Relative importance of each predictor for adherence sorted by group. BADS: Behavioral Activation for Depression Scale–Short Form; BMI: body mass index; CAQ: Cardiac Anxiety Questionnaire; EQ5D: European Quality of Life Questionnaire–Five Dimensions; HADS: Hospital Anxiety and Depression Scale ; MI: myocardial infarction; VAS: visual analog scale.



Discussion

Principal Findings

Our study tested and compared established and novel predictors for adherence to 14 weeks of therapist-supported iCBT using data from 90 MI-ANXDEP patients recruited from 25 hospitals in Sweden and randomized to treatment in the U-CARE Heart clinical trial. The time point of prediction was after completion of the first homework assignment, which therefore allowed the study of previously untested linguistic predictors extracted from actual written behavior together with previously established predictors. A robust machine learning procedure sifted out the most potent predictors for adherence assessed at the end of treatment, which were found to be self-assessed cardiac fear, sex, number of words, self-assessed general cardiac anxiety, average sentence length, and number of mutual words used.

Clinical Interpretation and Possible Implications

Both symptoms of general cardiac anxiety and specific cardiac fear were among the strongest predictors, and to the extent of symptom and mechanistic overlap, this corroborates previous findings that depression is associated with increased adherence to cardiac rehabilitation [43]. It is even more likely that cardiac anxiety, which is directly linked to the present treatment, would trigger activity more strongly than depression, given the respective symptomatology. Depression and anxiety are highly interconnected, which might explain the result from the cited study. Thus, patients reporting high levels of depression and patients reporting high levels of anxiety have acknowledged that they have a problem. Overall, it seems reasonable given the strength of the anxiety-based predictors that those MI-ANXDEP patients who are relatively less worried, in general and specifically regarding their heart, are less likely to adhere to treatment that specifically targets such symptoms. Our study also found that female sex was an important predictor for adherence, in line with pooled iCBT trial data confirming that males have a higher drop-out rate from Web-based interventions for depression [20]. Although not interchangeable, drop out is reasonably related to poor adherence.

On the other hand, our findings do not replicate other previously identified predictors for adherence to iCBT such as education and age [12,20], possibly due to the relatively old MI-ANXDEP patient population or the differing recruitment procedure in this study relative to the bulk of previous iCBT studies. Neither was alcohol a predictor, which might be due to a generally low level of problem drinking in the study sample. Although the U-CARE Heart inclusion had relatively high ecological validity compared to other iCBT studies, our patients were still selected, excluding, for instance, those with suicidal tendencies. Moreover, the weak predictive power of depression as gauged by the Hospital Anxiety and Depression Scale (HADS), especially compared to symptoms of anxiety and their strong predictive power, is puzzling. This may be due to exclusion of severe depressive symptoms on the basis of suicide risk, whereas no such screening was applied for very high anxious symptomatology. With that said, HADS anxiety was not a useful predictor, possibly suggesting psychometric shortcomings of the particular HADS scale. Consequently, the more cardiospecific anxiety scale CAQ seems more relevant for adherence in MI-ANXDEP patients. Furthermore, alternative ongoing treatment external to the trial (eg, psychoactive medication and third-party counseling) was not predictive of adherence to iCBT. Important to note is that there were no restrictions on patients seeking additional external treatment available from the relatively well-developed Swedish health care system. This could perhaps explain the null finding through the principle of homeostasis applied to symptom severity and sought treatment. In a relatively free and rich society, particularly severe symptomatology should be compensated for by such patients seeking and receiving multimodal treatment as needed. If so, these factors might cancel each other out with respect to both the need for and adherence to iCBT.

We also discovered that novel linguistic predictors based on written verbal responses predicted adherence. The number of words may be a proxy for verbal fluency and degree of patient effort in therapy, and the number of mutual words might be a proxy for the degree of therapeutic alliance, which in part corroborates previous research on therapeutic alliance and other interlinked concepts that promote adherence to iCBT [12,19,20,24,43]. Together with previously known predictors, these linguistic predictors may enable improved risk stratification regarding which patients will likely adhere to treatment. This suggests a largely unexplored route for future clinical research seeking to lower iCBT treatment failure and might lead to further tailoring of limited therapeutic resources for augmenting cost-effectiveness and lowering human suffering in clinical care.

Although more work is arguably needed, the data collection, preprocessing, and analysis of written responses can be automated to a considerable degree so the current lack of off-the-shelf clinical utility might not be a future obstacle. An automated tool for predicting adherence can be constructed and then possibly used as a decision support tool by the clinician. Moreover, the tool could also determine the risk of low adherence in patients, which could possibly inform the tailoring of treatment for the MI-ANXDEP patient more objectivity and accurately compared to the guesswork and crude cutoffs often applied to counter low adherence in clinical research and care today. So-called artificial intelligence and the related supervised machine learning applications that are now being rapidly researched and implemented broadly would likely also be of benefit to better solve the clinically relevant problem of predicting adherence to internet-delivered treatments.

Limitations and Strengths

A limitation of this study is the sample size. Although the present U-CARE Heart study is the largest iCBT trial for MI-ANXDEP patients to date, it provides limited reliability estimates. The sample is too small to subdivide for more detailed analyses of those exclusively depressed or anxious. Within the limits of the present sample size not allowing for an external validation data set, the generalizability of findings is, however, quite good given that (a) applied predictive modeling procedure was robustly cross-validated, (b) national coverage was very good with recruitment from 25 hospitals, and (c) patients were recruited very similarly to routine clinical care.

Although we used expert content knowledge to select predictors and tested a range of common and domain specific predictors, there was still the possibility for using other predictors. This might explain the room for improvement in terms of classification acuity. Given that we studied a whole new class of predictors consisting of actual written behavior selected by domain experts, this study adds further novelty in that manner. The confirmation of some previously known predictors for adherence to psychotherapy with scarcely studied but very common MI-ANXDEP patients indicates potential clinical utility with MI-ANXDEP patients. The study was conducted in Sweden, and we cannot readily extrapolate our findings beyond our national and linguistic borders. The MI-ANXDEP population is also a distinct subgroup of MI patients, and the iCBT intervention is specifically tailored to these patients. Hence, replication outside of Sweden with different patients and for other psychotherapeutic treatments would be valuable.

There was also the limitation of operationalizing the outcome. This can be done in several ways, with the most strict adherence definition being to complete all treatment modules [44]. However, since the U-CARE Heart trial had particularly high ecological validity but suffered from generally low adherence [13], this cutoff definition of adherence automatically had to be low to be able to model adherence since the moderate sample size inhibited us from modeling unbalanced classes. Defining adherence as those patients continuing treatment beyond the first two standardized modules is also arguably more clinically relevant on qualitative grounds compared to an arbitrary percentage cutoff. Considering clinical needs and data availability, the patients were selected on completion of the initial standardized homework module—the optimal time to predict treatment adherence if one wants to also use linguistic predictors derived from written treatment response to make early in-treatment prediction of treatment adherence. There are also qualitative approaches to investigate adherence to iCBT [25] that might augment our understanding of adherence if combined with the current data-driven approach. Furthermore, the purpose of studying linguistic predictors automatically excluded 27 patients who were randomized to treatment but did not complete any homework assignment. For obvious reasons, our prediction model cannot generalize to these patients, yet it seems likely that prediction accuracy would theoretically be higher if including these patients because they constitute extreme cases of low adherence.

Conclusions

For developing and testing effective iCBT interventions, investigating factors that predict adherence is important. Using a supervised machine learning approach, adherence to iCBT treatment in a multicenter trial for MI-ANXDEP patient was best predicted by a diverse set of predictors. The most potent predictors also included novel linguistic predictors from written patient behavior at the start of treatment. Our findings may improve the tailoring of iCBT for these high-risk patients. Future research should also investigate possible causal mechanisms and determine if these findings replicate outside of Sweden, in larger samples, and for other patient groups that might benefit from iCBT.

Acknowledgments

We are grateful to the U-CARE Heart patients. This study was supported by the Swedish Research Council for Health, Working Life, and Welfare (2014-4947), the Vårdal Foundation (2014-0114), and the strategic research program U-CARE (2009-1093).

Authors' Contributions

JW, EMGO, and EG designed the study. JW, EMGO, EG, CH, GM, FN, and LvE interpreted the findings, critically revised the manuscript, and approved its final form and submission. EG, JW, FN, and EMGO preprocessed data. JW analyzed data and drafted the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental material.
[PDF File (Adobe PDF File), 61 KB - jmir_v20i10e10754_app1.pdf]

Multimedia Appendix 2
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 677 KB - jmir_v20i10e10754_app2.pdf]

References

1. GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2015 Jan 10;385(9963):117-171. [doi: 10.1016/S0140-6736(14)61682-2] [Medline: 25530442]
2. Piepoli MF, Corrà U, Dendale P, Frederix I, Prescott E, Schmid JP, et al. Challenges in secondary prevention after acute myocardial infarction: a call for action. Eur J Prev Cardiol 2016 Dec;23(18):1994-2006. [doi: 10.1177/2047487316663873] [Medline: 27600690]
3. Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, Lanas F, INTERHEART Study Investigators. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. Lancet 2004;364(9438):937-952. [doi: 10.1016/S0140-6736(04)17018-9] [Medline: 15364185]
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. 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]
6. 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]
7. 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]
8. 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]
9. Piepoli MF, Hoes AW, Agewall S, Albus C, Brotons C, Catapano AL, ESC Scientific Document Group. 2016 European guidelines on cardiovascular disease prevention in clinical practice: the Sixth Joint Task Force of the European Society of Cardiology and other societies on cardiovascular disease prevention in clinical practice. Eur Heart J 2016 Dec 01;37(29):2315-2381 [FREE Full text] [doi: 10.1093/eurheartj/ehw106] [Medline: 27222591]

10. Olthuis JV, Watt MC, Bailey K, Hayden JA, Stewart SH. Therapist-supported Internet cognitive behavioural therapy for anxiety disorders in adults. Cochrane Database Syst Rev 2016 Mar 12;3:CD011565. [doi: 10.1002/14651858.CD011565.pub2] [Medline: 26968204]
11. Hedman E, Ljótsson B, Lindefors N. Cognitive behavior therapy via the Internet: a systematic review of applications, clinical efficacy and cost-effectiveness. Expert Rev Pharmacoecon Outcomes Res 2012 Dec;12(6):745-764. [doi: 10.1586/erp.12.67] [Medline: 23252357]
12. Alfonsson S. Treatment Adherence in Internet-Based CBT: The Effects of Presentation, Support and Motivation [doctoral thesis]. Acta Universitatis Upsaliensis: Uppsala University; 2016.
13. Norlund F, Wallin E, Olsson EMG, Wallert J, Burell G, von Essen L, et al. 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 Mar 08;20(3):e88 [FREE Full text] [doi: 10.2196/jmir.9710] [Medline: 29519777]
14. Norlund F, Olsson EMG, 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]
15. Waller R, Gilbody S. Barriers to the uptake of computerized cognitive behavioural therapy: a systematic review of the quantitative and qualitative evidence. Psychol Med 2009 May;39(5):705-712. [doi: 10.1017/S0033291708004224] [Medline: 18812006]
16. Woodford J, Farrand P, Bessant M, Williams C. Recruitment into a guided internet based CBT (iCBT) intervention for depression: lesson learnt from the failure of a prevalence recruitment strategy. Contemp Clin Trials 2011 Sep;32(5):641-648. [doi: 10.1016/j.cct.2011.04.013] [Medline: 21570485]
17. Andersson G, Hedman E. Effectiveness of guided internet-based cognitive behavior therapy in regular clinical settings. Verhaltenstherapie 2013;23(3):140-148. [doi: 10.1159/000354779]
18. Kolandaivelu K, Leiden BB, O'Gara PT, Bhatt DL. Non-adherence to cardiovascular medications. Eur Heart J 2014 Dec 07;35(46):3267-3276. [doi: 10.1093/eurheartj/ehu364] [Medline: 25265973]
19. El Alaoui S, Ljótsson B, Hedman E, Kaldo V, Andersson E, Rück C, et al. Predictors of symptomatic change and adherence in Internet-based cognitive behaviour therapy for social anxiety disorder in routine psychiatric care. PLoS One 2015;10(4):e0124258 [FREE Full text] [doi: 10.1371/journal.pone.0124258] [Medline: 25893687]
20. Karyotaki E, Kleiboer A, Smit F, Turner DT, Pastor AM, Andersson G, et al. Predictors of treatment dropout in self-guided web-based interventions for depression: an 'individual patient data' meta-analysis. Psychol Med 2015 Oct;45(13):2717-2726. [doi: 10.1017/S0033291715000665] [Medline: 25881626]
21. Pennebaker JW, Mehl MR, Niederhoffer KG. Psychological aspects of natural language. use: our words, our selves. Annu Rev Psychol 2003;54:547-577. [doi: 10.1146/annurev.psych.54.101601.145041] [Medline: 12185209]
22. Dirkse D, Hadjistavropoulos HD, Hesser H, Barak A. Linguistic analysis of communication in therapist-assisted internet-delivered cognitive behavior therapy for generalized anxiety disorder. Cogn Behav Ther 2015 Sep;44(1):21-32. [doi: 10.1080/16506073.2014.952773] [Medline: 25244051]
23. Zinken J, Zinken K, Wilson JC, Butler L, Skinner T. Analysis of syntax and word use to predict successful participation in guided self-help for anxiety and depression. Psychiatry Res 2010 Sep 30;179(2):181-186. [doi: 10.1016/j.psychres.2010.04.011] [Medline: 20483481]
24. Van der Zanden R, Curie K, Van Londen M, Kramer J, Steen G, Cuijpers P. Web-based depression treatment: associations of clients' word use with adherence and outcome. J Affect Disord 2014 May;160:10-13 [FREE Full text] [doi: 10.1016/j.jad.2014.01.005] [Medline: 24709016]
25. 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 Mar 16;20(3):e87 [FREE Full text] [doi: 10.2196/jmir.9690] [Medline: 29549067]
26. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd Edition. New York: Springer; 2009.
27. 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]
28. Vaglio J, Conard M, Poston WS, O'Keefe J, Haddock CK, House J, et al. Testing the performance of the ENRICHD Social Support Instrument in cardiac patients. Health Qual Life Outcomes 2004 May 13;2:24 [FREE Full text] [doi: 10.1186/1477-7525-2-24] [Medline: 15142277]
29. 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]
30. 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]
31. 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]

32. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [Medline: 6880820]
33. Beretta L, Santaniello A. Nearest neighbor imputation algorithms: a critical evaluation. BMC Med Inform Decis Mak 2016 Dec 25;16 Suppl 3:74 [FREE Full text] [doi: 10.1186/s12911-016-0318-z] [Medline: 27454392]
34. Jönsson P, Wohlin C. An evaluation of k nearest neighbour imputation using Likert data. METRICS '04 Proceedings of the Software Metrics, 10th International Symposium 2004:108-118.
35. Kowarik A, Templ M. Imputation with R package VIM. J Stat Softw 2016;74(7):1-16.
36. Breiman L. Random forests. Machine Learn 2001;45:5-32.
37. Breiman L, Friedman J, Olshen R, Stone C. Classification and Regression Trees. Belmont: Wadsworth Books; 1984.
38. Wallert J, Tomasoni M, Madison G, Held C. Predicting two-year survival versus non-survival after first myocardial infarction using machine learning and Swedish national register data. BMC Med Inform Decis Mak 2017 Jul 05;17(1):99 [FREE Full text] [doi: 10.1186/s12911-017-0500-y] [Medline: 28679442]
39. Kuhn M. Building predictive models in R using the Caret package. J Stat Soft 2008;28(5):1-26. [doi: 10.18637/jss.v028.i05]
40. Anthony L. AntConc version 3.4.4. Tokyo: Waseda University; 2016.
41. Östling R. Stagger: an open-source part of speech tagger for Swedish. Northern Eur J Lang Technol 2013 Sep 16;3:1-18. [doi: 10.3384/nejlt.2000-1533.1331]
42. R: A Language and Environment for Statistical Computing. Vienna: Foundation for Statistical Computing; 2015.
43. Zullo MD, Gathright EC, Dolansky MA, Josephson RA, Cheruvu VK, Hughes JW. Influence of depression on utilization of cardiac rehabilitation postmyocardial infarction: a study of 158,991 Medicare beneficiaries. J Cardiopulm Rehabil Prev 2017 Jan;37(1):22-29 [FREE Full text] [doi: 10.1097/HCR.0000000000000222] [Medline: 27755259]
44. 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]

Abbreviations

CAQ: Cardiac Anxiety Questionnaire
CBT: cognitive behavioral therapy
CVD: cardiovascular disease
HADS: Hospital Anxiety and Depression Scale
iCBT: internet-based cognitive behavioral therapy
k-NN: k nearest neighbor
MI: myocardial infarction
MI-ANXDEP: myocardial infarction with comorbid symptoms of depression, anxiety, or depression and anxiety
MAR: missing at random
MCAR: missing completely at random
RFE: recursive feature elimination
SBP: systolic blood pressure
U-CARE: Uppsala University Psychosocial Care Programme

Edited by G Eysenbach; submitted 11.04.18; peer-reviewed by C Păsărelu, F Lenhard, M Gonzalez, R Gallagher; comments to author 02.05.18; revised version received 20.06.18; accepted 05.07.18; published 10.10.18.

Please cite as:
Wallert J, Gustafson E, Held C, Madison G, Norlund F, von Essen L, Olsson EMG
Predicting Adherence to Internet-Delivered Psychotherapy for Symptoms of Depression and Anxiety After Myocardial Infarction: Machine Learning Insights From the U-CARE Heart Randomized Controlled Trial
J Med Internet Res 2018;20(10):e10754
URL: http://www.jmir.org/2018/10/e10754/
doi:10.2196/10754
PMID:30305255

Original Paper

Combining Technology and Research to Prevent Scald Injuries (the Cool Runnings Intervention): Randomized Controlled Trial

Jacqueline Burgess[1,2,3*], BComm (Hons); Kerrianne Watt[4*], BSc, PhD; Roy M Kimble[2,5*], MD-Qld, MBChB, FRACS; Cate M Cameron[6,7*], BSocWk (Hons), MPH, PhD

[1]Centre for Children's Burns and Trauma Research, University of Queensland, Brisbane, Australia

[2]Pegg Leditschke Children’s Burns Centre, Lady Cilento Childen’s Hospital, Brisbane, Australia

[3]Wound Management Innovation Cooperative Research Centre, Brisbane, Australia

[4]College of Public Health, Medical and Veterinary Sciences, James Cook University, Townsville, Australia

[5]Department of Paediatric Surgery, Urology Burns & Trauma Unit, Lady Cilento Children's Hospital, Brisbane, Australia

[6]Jamieson Trauma Institute, Royal Brisbane & Women’s Hospital, Metro North Hospital and Health Services District, Brisbane, Australia

[7]School of Public Health and Social Work, Queensland University of Technology, Brisbane, Australia

[*]all authors contributed equally

Corresponding Author:
Jacqueline Burgess, BComm (Hons)
Centre for Children's Burns and Trauma Research
University of Queensland
Child Health Research Centre
62 Graham Street
Brisbane, 4021
Australia
Phone: 61 407789078
Email: jacquii@uq.edu.au

Abstract

Background: New technologies, internet accessibility, social media, and increased smartphone ownership provide new opportunities for health researchers to communicate and engage target audiences. An innovative burn prevention intervention was developed using these channels.

Objective: The aim of this study was to evaluate the efficacy of Cool Runnings, an app-based intervention to increase knowledge of childhood burn risk (specifically hot beverage scalds) and correct burn first aid among mothers of young children.

Methods: This was a 2-group, parallel, single-blinded randomized controlled trial (RCT). Participants were women aged 18 years and above, living in Queensland, Australia, with at least 1 child aged 5-12 months at time of enrollment. The primary outcome measures were change in knowledge about risk of burns and correct burn first aid assessed via 2 methods: (1) overall score and (2) categorized as adequate (score=4) versus inadequate (score<4). Efficacy of gamification techniques was also assessed.

Results: In total, 498 participants were recruited via social media and enrolled. At the 6-month follow-up, 244 participants completed the posttest questionnaire. Attrition rates in both groups were similar. Participants who remained in the study did not differ from those lost to follow-up on any characteristics except education level. Although similar at baseline, intervention group participants achieved significantly greater improvement in overall knowledge posttest than control group participants on both primary outcome measures (overall knowledge intervention: mean [SD] of overall knowledge 2.68 [SD 1.00] for intervention vs 2.13 [SD 1.03] for control; 20.7% [25/121] adequate in intervention vs 7.3% [2/123] in control). Consequently, the number needed to treat was 7.46. Logistic regression showed participants exposed to the highest level of disadvantage had 7.3 times higher odds of improved overall knowledge scores than participants in other levels of disadvantage. There were also significant correlations between gamification techniques and knowledge change (*P*<.001). In addition, odds of knowledge improvement between baseline and 6-month follow-up was higher in participants with low-moderate app activity compared with no app activity (odds ratio [OR] 8.59, 95% CI 2.9-25.02) and much higher in participants with high app activity (OR 18.26, 95% CI 7.1-46.8).

Conclusions: Despite substantial loss to follow-up, this RCT demonstrates the Cool Runnings app was an effective intervention for improving knowledge about risks of hot beverage scalds and burn first aid in mothers of young children. The benefits of combining gamification elements in the intervention were also highlighted. Given the low cost and large reach of smartphone apps to deliver content to and engage with targeted populations, the results from this RCT provide important information on how smartphone apps can be used for widespread injury prevention campaigns and public health campaigns generally.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12616000019404; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=369745&showOriginal=true&isReview=true (Archived by WebCite at http://www.webcitation.org/72b1E8gTW)

(J Med Internet Res 2018;20(10):e10361) doi:10.2196/10361

KEYWORDS

burns; infant; child; mobile apps; gamification; injury; prevention; parent

Introduction

App Technology

Advances in technology, expansion of internet access, and increased mobile phone ownership globally have led to a new channel for disseminating health information and engaging with large or specific populations. With the popularity of smartphones, there has been a proliferation of smartphone apps—6 million in the 2 leading app stores (Google Play: 2.8 million, Apple app store: 2.2 million) [1]. Of these, 259,000 are health-related apps [2]. Increasingly, apps are being used by health agencies and researchers to gather and present information to study participants and the general public. There is a growing body of evidence showing the successful use of smartphone apps to encourage healthy habits such as increasing physical activity [3] and promoting weight loss [4], managing chronic diseases [5,6], and delivering mental health programs [7]. One area that has not yet been studied is the use of this technology in injury prevention.

Childhood Burns

Childhood burns are serious injuries that can result in substantial pain and suffering and lead to life-long scarring and surgical procedures as the child grows. The physical, emotional, and financial burden to the child and family can be significant [8,9]. The leading cause of childhood burns in developed countries is hot drink scalds [10-13]. In Australia, hot drink scalds account for 18% of all childhood burns [14,15]. This injury peaks in children aged 6 to 18 months, usually occurs in the child's home, and is witnessed by the parent or supervising adult [13-17]. Given these facts, an app-based prevention intervention was developed to target mothers with children aged 5 to 12 months about risks of hot drink scalds, as well as the correct first-aid treatment to apply, should a burn occur.

Cool Runnings

On the basis of the Health Belief Model [18], the aim of the Cool Runnings randomized controlled trial (RCT) was to assess the impact of a contemporary app-based public health campaign using gamification on knowledge about child burns. Specifically, the aim of the intervention was to increase knowledge of the primary carer about the severity and frequency of hot drink scalds, provide them with developmental-stage messages on how to protect their child and intervene, and finally the correct burn first-aid treatment to apply should a burn occur. The 2 aims of this study were therefore to (1) assess change in knowledge from baseline to follow-up in the intervention group compared with the control group and (2) investigate the impact of level of app engagement on change in knowledge from baseline to follow-up.

Methods

Study Design

This study was a 2-group, parallel, single-blinded RCT of an app-based prevention and first-aid education intervention for burns. This study was registered with the Australian and New Zealand Clinical Trials Registry (ACTRN12616000019404). The full study protocol has been published previously [19].

Study Setting

Participants from Queensland, Australia, were recruited. Eligibility criteria were females aged 18 years and above; who resided in Queensland, Australia; and had at least 1 child aged 5 to 12 months at enrollment. Ownership of a smartphone was required for intervention delivery. Study duration was 6 months.

Recruitment

Participants were recruited via online social media advertisements, specifically through Facebook and Instagram, between January 2016 and February 2016. A detailed description of the recruitment process for this study has been published previously [20].

Randomization

Computerized sequence generation was used to randomize participants. Randomization was stratified by maternal age (18-28 years and 29+ years) based on the mean national maternal age [21].

Blinding

Participants were blinded to their allocation group (the terms *blue group* and *green group* were used). Study investigators assessed the outcome data collected in pre- and postquestionnaires in a blinded format. However, blinding was not possible for analyzing the results of gamification techniques, as they only applied to the intervention group.

Intervention

Participants in the intervention group were compared with those in the control group. Both groups accessed an app throughout the study, but in the intervention group, gamification techniques were incorporated into the app. The control group accessed a slightly different version of the app with no gamification.

Intervention Group

During the 6-month intervention, participants allocated to the intervention group received 9 intervention messages via the app related to risks of hot beverage scalds, risks of developmental stage–based burns, and burn first-aid treatment (illustrated in Figure 1). These messages were provided in a variety of mediums (infographics, 30-second videos, and motion graphics) at 3-week intervals. In between these messages, participants were given opportunities to engage with the app through activities such as answering pop quizzes and completing missions (such as photo uploads) that reinforced each of the intervention message themes. Gamification techniques were used to keep participants engaged and active on the app. Each time participants viewed a message, correctly answered a quiz question, or uploaded a photo, they were rewarded with points. Accrued points were displayed on weekly leaderboards in the app, and once a certain number of points were reached, they could be redeemed for rewards, such as shopping and movie vouchers.

Control Group

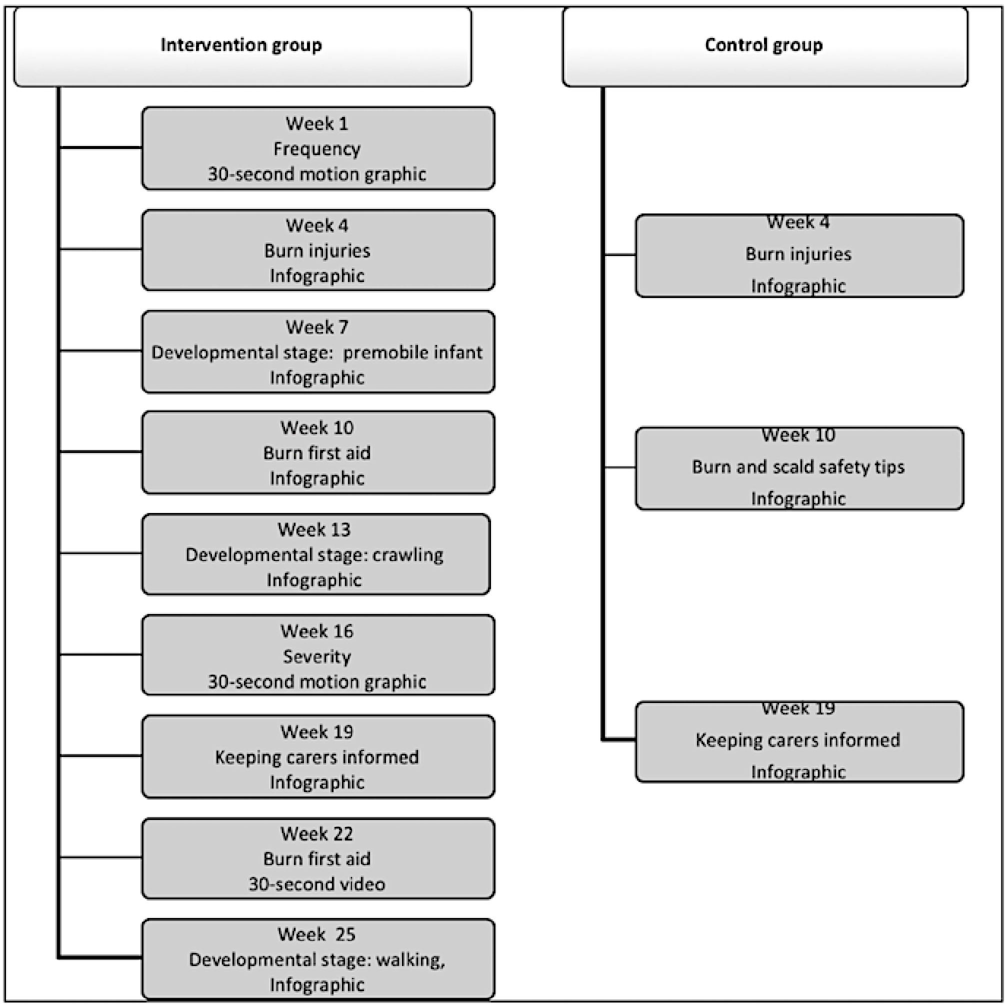
This group accessed a slightly different app interface; no gamification techniques were used with this group. Participants from the control group only received 3 messages during the 6-month intervention. These messages were infographics, and there were no opportunities for participants to engage with the material.

Data Collection

Baseline and 6-month follow-up questionnaires were completed by participants in the intervention and control groups. The baseline questionnaire included demographic factors (such as education level, age of youngest child, number of children, marital status, and smoking status). Place of residence postcodes were also collected and later recoded using the Accessibility/Remoteness Index of Australia (ARIA) 2011 data [22], and the Socio-Economic Indexes for Areas (SEIFA) [23] as measures to broadly assess socioeconomic status (SES). The SEIFA data were based on aggregate area-level SES disadvantage indicators and were categorized into quintiles (1=most disadvantaged and 5=least disadvantaged). The ARIA is a measure of geographical remoteness, categorized as urban, periurban, and remote.

The questionnaires also included the extent of hot beverage scald risk awareness (2 questions) and burn first aid knowledge (2 questions). Full baseline data from this study are described elsewhere [24].

Figure 1. Content calendar for the Cool Runnings intervention (source: Burgess et al [19]).



The 6-month follow-up questionnaire repeated the questions relating to risks of hot beverage scalds and burn first aid knowledge. Participant engagement with the app, including number of app opens, content views, and gamification activities by participants in the intervention group were recorded by the app.

Primary Outcome Measures

The primary outcome for this study was change in knowledge based on a 4-point knowledge score measured by 3 components:

1. Knowledge of correct burn first aid,
2. Knowledge of the main cause of burns or scalds in children aged 0 to 15 years, and
3. Knowledge of the main age group at risk of these burns or scalds.

Two questions were used to determine burn first aid knowledge: an open-ended question ("What is the recommended first aid treatment for a burn or scald?") and a multiple-choice question regarding duration for applying cool running water. Correct first aid knowledge was defined as cool running water for 20 min based on clinical evidence of benefit [25]. Participants who responded with "20-minutes cool running water" to the open-ended question were allocated 2 points. Participants whose answer involved cool running water but mentioned an incorrect or no time and who then responded to the multiple-choice question on duration correctly were awarded 1.5 points; only 1 point was awarded if the subsequent multiple-choice question on duration was incorrect. All other responses (such as flour and ice), including "don't know," were allocated 0 points.

Knowledge of main cause of burns or scalds in children aged 0 to 15 years was assessed via a multiple-choice question. Anything other than hot drink scalds (1 point) was coded incorrect (0 points). Knowledge of main age group at risk of these burns or scalds was also assessed via multiple-choice questions. Anything other than 0 to 2 years (1 point) was coded incorrect (0 points).

One final overall knowledge score was then computed, which combined the responses to the main cause of burns, age group most at risk, and burn first aid knowledge, to yield a total possible score of 4. The change in total score on *overall knowledge* between baseline and postintervention was calculated.

The score was also recoded to a binary variable. Participants who received a score of 4 were coded as *adequate knowledge*, and all other participants were coded as *inadequate knowledge*. Participants whose responses moved from *inadequate* to *adequate* at 6-month follow-up were categorized as *improved*. All other participants were coded as *no improvement*.

In addition, burn first aid knowledge was categorized into a binary variable: adequate (cool running water for 20 min in the open-ended question) versus inadequate (any other response). Participants whose responses moved from *inadequate* to *adequate* at 6-month follow-up were categorized as *improved*. All other participants were coded as *no improvement*.

Gamification and App Engagement

The app accessed by the intervention group incorporated gamification techniques. Participants in the intervention group were encouraged to earn points by viewing content, completing pop quizzes, and uploading photos as part of weekly challenges. Weekly leaderboard and challenge winners were awarded additional bonus points. Level of engagement was measured across following 4 dimensions: the frequency of opening the app (app views: intervention and control groups), frequency of content views (intervention and control), number of pop quiz completions (intervention), and participation in photo-sharing activities (intervention). These numerical and continuous variables were each then categorized into none, low-moderate, and high. High engagement occurred if the participant engaged with at least 2 out of 3 of the available engagement opportunities for that element (eg, there were 6 opportunities to upload photos, so participants were required to upload 4 photos to be coded as high engagement).

One final measure of app engagement was then derived for all participants (intervention and control) as follows: no engagement on any of the 4 elements other than opening the app; high engagement (viewing content 4 or more times, sharing photos 4 or more times, or completing quizzes 9 or more times); and low-moderate engagement (any other level of engagement).

Data Analysis

All statistical analyses were conducted using SPSS version 24 (IBM Corporation, Armonk, NY, USA). Descriptive analyses were completed to determine whether there were any between-group differences (intervention vs control) at baseline on demographic characteristics and the primary outcome measure. Chi-square tests were used for categorical variables, and independent sample *t* tests were used for numerical variables [26]. Specifically, an independent sample *t* test was used to assess between-group differences on change in overall knowledge score at 6-month follow-up as a function of the intervention. A chi-square test was performed to determine whether the proportion of participants with improved knowledge differed between intervention and control groups. Event rate of improved overall knowledge (all 4 responses correct) was also calculated for the intervention and control groups. Subsequently, the number needed to treat (NNT) was calculated. Correlations were performed to determine whether the 4 separate elements of app engagement were related to each other. Alpha of .05 was used in the interpretation of all descriptive analyses. Univariate logistic regression analyses were conducted to determine whether there were any significant independent predictors of knowledge improvement (no improvement vs improvement). Potential predictor variables were intervention status (intervention vs control) and demographic variables (education level, age of youngest child, age of respondent, number of children, marital status, smoking status, ARIA category, SES as measured through SEIFA, and first-time mother).

Any variables where $P<.20$ was obtained in univariate logistic regression analyses were then entered into 1 adjusted model. If a variable was not significantly associated with the outcome in the multivariate model, it was removed and the impact on all remaining variables was assessed. If the odds ratio (OR) for any other variables in the model changed more than 10%, the variable was retained in the model as a potential confounder. If not, it was removed. This process was repeated until there were no variables with $P>.05$ in the model or removing the nonsignificant variables from the model did not create changes of greater than 10% to the ORs of variables remaining in the model.

To investigate the impact of level of app engagement on change in knowledge from baseline to follow-up, univariate analyses were first completed using the 4 numerical measures of engagement for the intervention group only (frequency of app views, frequency of content views, number of pop quiz completions, and number of times participated in photo-sharing activities). Afterward, for all participants (intervention and control), univariate analyses were completed on the final composite measure (no engagement, moderate engagement, and high engagement). Subsequently, an additional multivariate analysis was completed using this composite measure of app engagement as one of the predictor variables, instead of intervention status—the same demographic variables described above were used—and the same process followed. Analyses completed on any follow-up data were conducted on a *per protocol* basis.

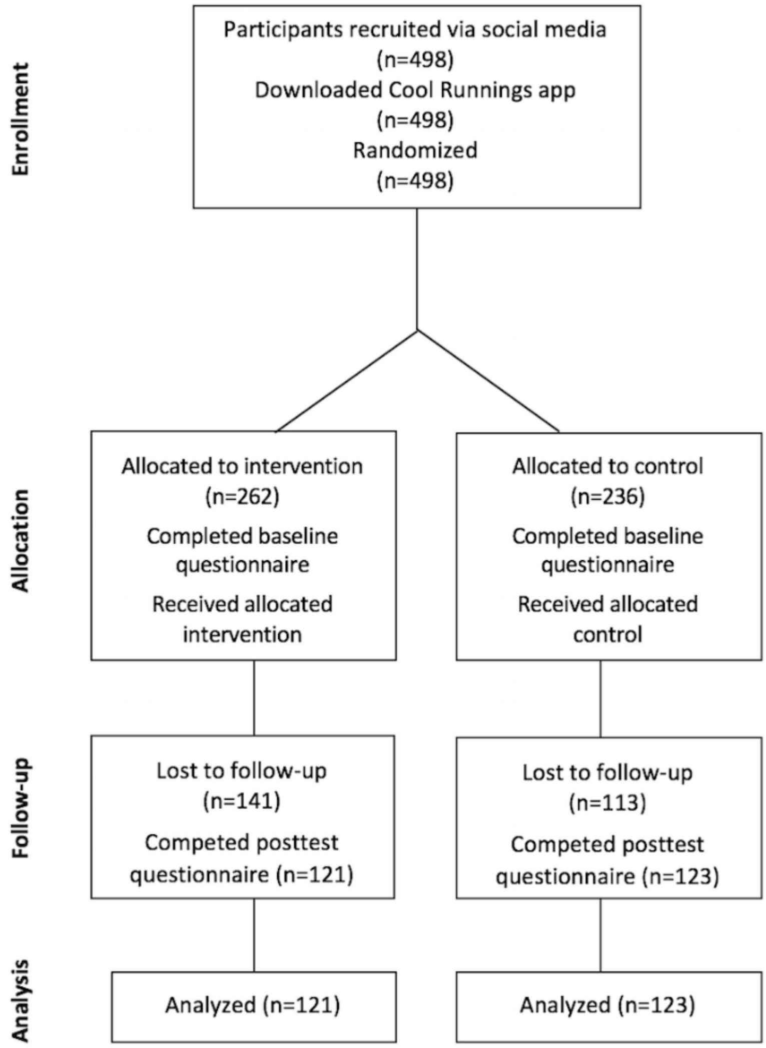
Results

Figure 2 illustrates the flow of participants through the study. A total of 498 participants were enrolled in the Cool Runnings study: 262 in the intervention group and 236 in the control group. After the 6-month intervention, 121 intervention participants (121/262, 46.1%) and 123 control participants (123/236, 52.1%) completed the posttest questionnaire.

Trial Retention

The trial experienced 51% attrition overall. Attrition rates in both groups were similar: (intervention: 141/262, 53.8%; control: 113/236, 47.8%). Participants who remained in the study did not differ from those who were lost to follow-up on any baseline characteristics except for education level. A higher proportion of participants who remained in the study had a university degree (28.7%; n=70) than those who were lost to follow-up (16.5%; n=42; χ^2_4=15.8; P=.003). Mean overall knowledge was higher at baseline in participants (mean 2.06 [SD 0.87]) than in those who were lost to follow-up (mean 1.93 [SD 0.87]), but this difference was not significant (t_{490}=1.72; P=.09).

Figure 2. Flowchart of participants through each stage of the randomized controlled trial.



Within-group analyses were also conducted to see whether there were differences in participants who completed the study and those who were lost to follow-up. There was no difference between participants who completed the study and those who were lost to follow-up in relation to proportion with adequate overall knowledge (score of 4) versus inadequate (score<4) in the intervention group (*P*=.62) or in the control group (*P*=.99). However, among the participants allocated to the intervention group, overall knowledge at baseline was significantly higher in those who remained in the study (mean 2.12 [SD 0.84]) than in those who did not complete the study (mean 1.84 [SD 0.87]; t_{258}=2.64; *P*=.009). The remainder of the analyses were completed by *compliance only*.

Primary Outcome Measures

At baseline, there were no differences in any demographic or other sample characteristics between the intervention and control groups (*P*>.05; see Table 1 for actual *P* values). Importantly, there were no differences in the mean total knowledge score (*P*=.54), the proportion of participants who demonstrated adequate overall knowledge (*P*=.49), or in any of the dimensions comprising the score (burn risk knowledge [*P*=.93] and burn first aid knowledge [*P*=.57]; Table 2).

Changes in overall knowledge of participants between baseline and 6-month follow-up are shown in Table 2 and Figure 3. Although similar at baseline, intervention group participants achieved significantly greater improvement in overall knowledge posttest than control group participants (t_{240}=3.37; *P*<.001; Figure 3). *Event rate* of improved overall knowledge (change from inadequate at baseline to adequate at 6-month follow-up) was significantly higher in the intervention group (25/121, 20.7%) than in the control group (9/123, 7.3%) (χ^2_1=9.1; *P*=.003). Consequently, the NNT was 7.46. That is, 8 people needed to be exposed to this intervention to improve inadequate overall knowledge to adequate knowledge (ie, score of <4 to a score of 4) in 1 additional person. A sensitivity analysis was completed with respect to the event rate and NNT. First, the event rate was recalculated assuming that all participants who were lost to follow-up did not improve their score (ie, demonstrated inadequate knowledge at baseline and at follow-up). The event rate of improved overall knowledge was 9.5% in the intervention group and 3.8% in the control group. The NNT was 17.5. Next, the event rate was recalculated assuming that all participants who were lost to follow-up did improve their score from inadequate at baseline to adequate at follow-up (intervention: 63.35%; control: 51.69%). The NNT was 8.57.

Demographic Predictors of Overall Knowledge of Burns

Univariate logistic regressions indicated that the following variables were related to improvement in overall knowledge (from inadequate to adequate) between baseline and follow-up: being in the intervention group, age of respondent, SES as measured through SEIFA quintile, and remoteness (as measured by ARIA category). These variables were entered into 1 multivariate model ,and nonsignificant variables were removed one at a time, assessing the impact on remaining variables. In the final model, the only variables that were significantly associated with the improvement in overall knowledge were being allocated to the intervention (OR 3.3, 95% CI 1.4-7.7) and SES as measured through SEIFA. Specifically, odds of improving overall knowledge scores were higher in participants whose postcode indicated they were exposed to the highest level of disadvantage (OR 7.30, 95% CI 1.2-42.9) compared with participants exposed to the lowest levels of disadvantage (ie, highest advantage). Age and remoteness (as measured by ARIA category) were not significantly associated with improved knowledge; however, they were retained in the model because there was evidence of confounding (ie, removing these variables from the model changed the ORs of other variables in the model more than 10%).

Gamification and App Engagement

Gamification and app use activity (app views, content views, pop quiz completions, and photo sharing) were also measured. Participants in the intervention group earned points each time they viewed content, correctly answered pop quiz questions, and uploaded photos. Winners of the weekly photo mission won additional bonus points. In total, 58 participants in the intervention group accrued sufficient points to redeem their points for movie or supermarket vouchers; however, only 3 participants took advantage of this. The leaderboard (only available to the intervention group) showed that participants' points were viewed 535 times.

The mean number of app opens for the intervention group was 18.31 (SD 42.1; minimum 1; maximum 347; median 5.0; interquartile range [IQR] 13.5) and 5.03 (SD 5.28; minimum 1; maximum 28) for the control group (median 3.0; IQR 4.0). Overall, 1 participant from the intervention group opened the app a total of 347 times; however, the majority of participants opened the app 10 times or less in both the intervention group (69%) and control group (65%). Mean content views for the intervention group was 1.96 (SD 2.86; median 0.0, IQR 4.5) and 0.98 (SD 0.77) for the control group (median 0; IQR 0). The mean quiz completions for the intervention group was 2.45 (SD 4.33), the median was 0 (IQR 2.5), and the mean number of photos shared was 2.23 (SD 5.11; median 0; IQR 2.0). With respect to the composite measure of app engagement, 27.3% (33/121) of the intervention group and 1.6% (2/123) of control participants were categorized as high engagement and 51.2% (62/121) of the intervention group had no app engagement (62) versus 98.4% (121/123) of the control group.

Univariate logistic regression analyses (intervention group only) showed that each of the 4 (numerical) measures of engagement were significantly associated with improvement in overall knowledge from inadequate at baseline to adequate at 6-month follow-up (quiz total: OR 1.34, 95% CI 1.2-1.5; content: OR 1.49, 95% CI 1.3-1.7; app opens: OR 1.05, 95% CI 1.02-1.07; and photo uploads: OR 1.33, 95% CI 1.1-1.6). The 4 elements of app engagement were strongly correlated with each other (see Table 3).

Table 1. Demographic characteristics and knowledge of risks of scalds and first aid in intervention and control groups at baseline.

Characteristics	Intervention (n=121), n (%)	Control (n=123), n (%)	*P* value
Age of participant (in years)			.56
18-24	19 (15.7)	20 (16.3)	
25-29	34 (28.1)	43 (35.0)	
30-34	46 (38.0)	44 (35.8)	
35+	22 (18.2)	16 (13.0)	
Marital status			.21
Single	15 (12.4)	9 (7.3)	
Married	76 (62.8)	70 (56.9)	
De facto	28 (23.1)	40 (32.5)	
Separated or divorced	2 (1.7)	4 (3.3)	
Highest education level			.62
Less than year 12	13 (10.7)	22 (17.9)	
Year 12 completion	27 (22.3)	25 (20.3)	
Technical and further education certificate or advanced diploma	34 (28.1)	33 (26.8)	
University degree	37 (30.6)	33 (26.8)	
Postgraduate degree	10 (8.3)	10 (8.1)	
Current smoker			.70
Smoker	23 (19.0)	21 (17.1)	
Nonsmoker	98 (80.9)	102 (82.9)	
Country of birth			.41
Australia	104 (85.9)	101 (82.1)	
Other	17 (14.0)	22 (17.9)	
SEIFA[a]			.92
1 (most disadvantaged)	10 (8.3)	10 (8.2)	
2	20 (16.5)	16 (13.1)	
3	32 (26.4)	30 (24.6)	
4	46 (38)	51 (41.8)	
5 (least disadvantaged)	13 (10.7)	15 (12.3)	
ARIA[b]			.64
Urban (major cities)	62 (51.2)	62 (50.8)	
Periurban (inner or outer regional)	44 (36.4)	49 (40.2)	
Remote or very remote	15 (12.4)	11 (9.0)	
First-time mothers			.50
Yes	48 (39.7)	54 (43.9)	
No	73 (60.3)	69 (56.1)	
Number of children in the household			.25
1 child	53 (43.8)	63 (51.2)	
More than 1 child	68 (56.2)	60 (48.8)	

[a] SEIFA: Socioeconomic Index for Areas. SEIFA was used to estimate socioeconomic status in this study. Specifically, the Index of Relative Socioeconomic Advantage and Disadvantage. Higher deciles reflect higher relative advantage, and lower deciles reflect lower relative advantage. Deciles were reduced to 5 categories.

[b] ARIA: Accessibility/Remoteness Index of Australia. Location of usual residence was categorized using ARIA, developed by National Centre for the Social Applications of Geographic Information Systems. Each geographical area was allocated a score between 0 and 15, based on the (road) distance

to nearby towns that provide services. Scores were then allocated to the following categories (Office of Economic and Statistical Research Queensland, 2011): urban (major city: 0.0-0.2); periurban (inner regional: 0.2-2.4 and outer regional: 2.4-5.92); and remote (remote: 5.92-10.53; very remote: 10.53+).

Table 2. Change in overall knowledge and burn first aid at 6-month follow-up.

Knowledge metrics	Intervention (n=121)		Control (n=123)	
	Pretest[a]	Posttest	Pretest	Posttest
Overall knowledge, mean (SD)	2.11 (860)	2.68 (1.00)	2.04 (0.915)	2.13 (1.03)
Overall knowledge[b]—adequate, n (%)	3 (2.5)	27 (22.3)	5 (4.1)	12 (9.8)
Main cause of burns or scalds—correct, n (%)	40 (33.1)	72 (59.5)	44 (35.8)	57 (46.3)
Age group most at risk of burns or scalds—correct, n (%)	64 (52.9)	72 (59.5)	56 (45.5)	48 (39.0)
Burn first aid knowledge, n (%)				
Adequate (cool running water for 20 min, unprompted)	12 (9.9)	48 (39.7)	15 (12.4)	32 (26.0)
Inadequate (all other responses)	109 (90.1)	73 (60.3)	108 (87.6)	91 (73.9)

[a] Participants lost to follow-up are not included in the Pretest column.

[b] Proportion of participants who had *adequate* overall knowledge at baseline and 6-month follow-up are shown in the table. These rates are slightly different from the event rates that were calculated for improvement in overall knowledge between baseline and follow-up (25 [20.7%] participants in the intervention group improved overall knowledge between baseline and follow-up compared with 9 [7.3%] people in the control group). Only those participants who demonstrated improvement from inadequate knowledge at baseline to adequate knowledge at 6-month follow-up were included in the *improved knowledge* group).

Figure 3. Change in mean overall knowledge score between baseline and 6-month follow-up in intervention and control groups (error bars: 95% CI).

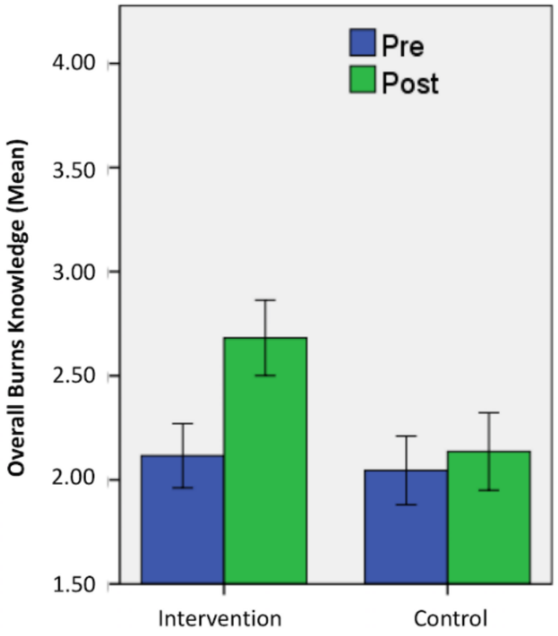


Table 3. Intercorrelations between measures of app engagement (N=244).

App activities	Change in knowledge	Quiz completion	Content view	App opens
Quiz	.48[a]	—	—	—
Content	.44[a]	.86[a]	—	—
App	.35[a]	.72[a]	.56[a]	—
Photo uploads	.40[a]	.83[a]	.65[a]	.76[a]

[a] *P*<.001.

Univariate logistic regression analyses (intervention and control groups) indicated that the composite measure of app engagement was associated with change in knowledge at 6-month follow-up. Odds of improved overall knowledge from inadequate to adequate were significantly higher in participants who demonstrated low-moderate app engagement (OR 8.59; 95% CI 2.9-25.02) and high app engagement (OR 18.26; 95% CI 7.1-46.8) than participants with no engagement. The composite measure of app engagement was entered into a multivariate logistic regression model with the variables previously identified in univariate analyses as significantly associated with the primary outcome measure (age of respondent, SES as measured through SEIFA quintile, and remoteness as measured by ARIA category). Nonsignificant variables were removed from the model one at a time and the impact on remaining variables was assessed. In the final model, the only variable that was significantly associated with improvement in overall knowledge was app engagement (low-moderate: OR 6.81; 95% CI 2.2-21.4 and high: OR 33.84; 95% CI 10.6-107.6). Age of respondent, remoteness (as measured by ARIA category), and SES (as measured through SEIFA) were not significantly associated with app engagement; however, they were retained in the model because there was evidence of confounding (ie, removing these variables from the model changed the ORs of other variables in the model more than 10%).

Discussion

Principal Findings

This RCT has demonstrated the Cool Runnings app to be an effective intervention for improving knowledge about risks of hot beverage scalds and burn first aid in mothers of young children. Only 8 people needed to be exposed to this intervention to improve inadequate overall knowledge to adequate knowledge in 1 additional person. Hot beverage scalds present a major pediatric public health issue that requires attention and prevention efforts, and this RCT details the implementation and evaluation of innovative methods and techniques to address this injury.

To our knowledge, this is the first study to evaluate an app-based delivery of injury prevention messages and the first study to gauge the efficacy of gamification in an injury prevention intervention. Given the low cost and large reach of smartphone apps to deliver content to and engage with targeted populations, results from this RCT provide important information on how smartphone apps can be used for widespread injury prevention campaigns. This study looked specifically at the use of this technology in a prevention campaign aimed at hot beverage scalds, the leading cause of childhood burn injuries. Although numerous studies have reported the high incidence of this injury, there is a paucity of interventions aimed at preventing them.

There have been few success stories when it comes to prevention campaigns for childhood burns [27], apart from those that have included passive approaches that have the benefit of legislative, engineering, and design support, such as hot water tempering valves and flame-retardant children's sleepwear [28-30]. Certainly, prevention campaigns for burns that focus solely on education have had little demonstrated success. Due to this and because education is the most likely strategy to be effective for hot beverage scald prevention (because of limited capacity for other approaches such as engineering or environmental approaches to work), a novel approach was followed to develop and implement the Cool Runnings intervention.

In addition to being allocated to the intervention group, SES (as measured by SEIFA through postcode of residence) was significantly inversely associated with improvement in knowledge score relative to baseline. This is encouraging, given the recognized disparities in burn incidence and first aid knowledge and use among those who are socioeconomically disadvantaged [31-34]. Some burn prevention campaigns have targeted these specific groups with mixed results [35,36]. The growing global ownership of smartphones and the promise of app-based technology may change this. In 2016, the average global ownership of smartphones was 81% (77% of US adults and 84% Australian adults) [37]. Lack of other significant predictors of increased burn knowledge may be interpreted as an indication of the success of the intervention across the target group. There was a broad sample of participants included in this study—older and younger primigravid mothers, with various levels of education, and from regional, rural, and remote locations. Participants were representative of the target population (women who delivered in Queensland in 2015) with regard to age, marital status, being a first-time mother, and country of birth [38].

Smartphone ownership goes beyond socioeconomic, racial, and ethnic boundaries, with a report by the Pew Institute [39] showing that more than half of most sociodemographic groups own a smartphone. Smartphones provide the opportunity to engage with people wherever they are and whenever it suits them to see a message. The use of smartphones to deliver information in a way that is interactive and engaging, rather than a one-way flow of static communication, is also compelling. App-based campaigns can cater to participants' different learning styles—whether it's visual, auditory, or kinesthetic—through the delivery of various message types, such as animations, videos, or infographics, and make it more appealing to a diverse audience.

This study also showed an association between change in knowledge and the gamification strategies used. Gamification takes the gaming principles of rewards, competition, and personalization to engage participants and motivate them toward preferred behaviors. Gamification is widely used in business to increase loyalty and create long-term engagement, but the evidence of its efficacy in changing health behaviors is still in its infancy. Gamification is commonly used in workplace health initiatives [40], and the trend of using gamification techniques and strategies in health-related apps is burgeoning. However, theoretical frameworks are still being developed, and there remains relatively little scientific literature as to its efficacy in improving health behavior outcomes [40,41]. The aim of incorporating gamification elements to health-related nongames is to improve user experience and engagement while increasing intrinsic motivators likely to result in the adoption of a behavior or knowledge change [42,43]. Extrinsic and intrinsic rewards were used in this intervention, with the potential to win financial rewards and prizes (extrinsic), and learn how to protect their child from injuries and/or keep children safe (intrinsic). In this study, participants in the intervention group were *gamed* into viewing content, uploading photos, and completing pop quizzes with the ability to earn points for each activity. These points accumulated and were displayed on the weekly leaderboard on the app. Although 58 participants accrued enough points to redeem them for tangible rewards (Aus $25 and Aus $50 movie or shopping vouchers), only 3 participants did so. Reasons for this may be the intangible rewards (such as leaderboard position and weekly photo winner badges) were of more perceived value, or the motivation may have been more competition-based rather than incentive-based. This finding shows potential for larger widespread campaigns in which providing ongoing tangible financial rewards would make scalability difficult. The results of this study suggest the inclusion of gamification strategies in injury prevention and public health campaigns could lead to improved results. Unfortunately, the use of gamification techniques did not appear to help with retention of participants. One reason for this may be the length of the intervention. It is difficult to know at what specific point during the intervention period participants dropped out. The lack of personal contact associated with an app-based intervention may also have contributed to the attrition rate. Future studies should test shorter intervention periods to determine optimal duration.

The loss to follow-up in this RCT provides some important information about the challenges associated with recruiting and retaining participants using this technology. Overall, the potential differences in the participants versus nonparticipants with respect to burn first aid knowledge and the differential loss to follow-up in the intervention group of those with lower knowledge indicates that further research is required regarding those with lower knowledge (who would most benefit from an intervention to improve knowledge), and how to engage or retain their interest and participation. Sensitivity analyses were conducted to further understand the potential impact of the loss to follow-up. When it was assumed that all participants who were lost to follow-up improved their knowledge score from inadequate to adequate, the NNT was 8.57. When it was assumed that all participants lost to follow-up did not improve their knowledge score, the NNT was 17.5 (compared with NNT of 7.46 calculated for the participants who were retained in the study).

Limitations

This study has several limitations. In both the control and intervention groups, there was a large loss to follow-up (48.9%). This loss to follow-up raises the potential for attrition bias. However, the attrition rate in both groups was similar (54% vs 48%), and participants did not differ significantly from those who were lost to follow-up on most of the measured characteristics. The exception was education (participants who remained in the study demonstrated a higher level of education). In addition, participants originally allocated to the intervention group who completed the study demonstrated a significantly higher baseline overall knowledge score than those who were lost to follow-up. This did not occur in the control group. Interestingly, there was no difference in the proportion of participants versus dropouts who demonstrated adequate versus inadequate knowledge. It is also acknowledged that there may have been differences between participants who remained in the study and those who were lost to follow-up that were not measured in the survey. Given the novelty of this intervention, and in particular within this context, we intentionally conducted a *per protocol* analyses to demonstrate efficacy of the app, and this may be considered a limitation of the analyses, although sensitivity analyses were conducted to further understand the potential impact of the loss to follow-up. The relatively small numbers involved in this study mean that the multivariate analyses on demographic variables associated with change in knowledge (especially when app engagement is considered) and the analyses on predictors of app engagement should be interpreted with caution.

Postcodes were used as a proxy for SES in this study. Hence, this measure may not have been representative of the individual-level SES of participants. It is possible that other factors external to study participation (such as enrollment in a first aid course) may have contributed to the observed change in outcome. However, this was an RCT, so the likelihood of this happening unequally between intervention and control groups is low and the potential impact on observed results is minimal.

Finally, it is important to acknowledge that a change in knowledge does not necessarily reflect a change in behavior. It was beyond the scope of this study to assess the impact of the RCT on behavior; however, an important next step would be to determine whether this app can affect behavior change in relation to burn first aid in young children with hot beverage scalds.

Conclusions

Despite the loss to follow-up, an app-based prevention intervention for burns appears to be an effective and appealing approach for targeting mothers of young children. These results have shown that only 8 people needed to be exposed to this intervention to improve inadequate overall knowledge to adequate knowledge (ie, score of less than 4 to a score of 4) in 1 additional person. The broad reach, low cost, and scalability of this medium could potentially be feasible for other injury prevention campaigns aimed at this population, particularly given the fact that children aged 0 to 4 years are most at risk for a number of injuries that occur in the home. Additional studies are needed to determine the optimal follow-up time for this type of intervention to offset the high attrition rate noted in this intervention.

Acknowledgments

JB would like to acknowledge the support of the Australian Government's Cooperative Research Center's Program. The authors would like to acknowledge iPug Pty Ltd for its expertise in developing the Cool Runnings app and guidance with gamification strategies used in this RCT.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT‐EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 1MB - jmir_v20i10e10361_app1.pdf]

References

1. Statista. Number of Facebook users worldwide 2008-2017 URL: https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/ [accessed 2017-06-09] [WebCite Cache ID 6r6BWbhrt]
2. Research2Guidance. mHealth App Developer Economics 2016 URL: https://research2guidance.com/product/mhealth-app-developer-economics-2016/ [accessed 2018-09-21] [WebCite Cache ID 72aS0yuKv]
3. Schoeppe S, Alley S, Van Lippevelde W, Bray N, Williams S, Duncan M, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
4. Flores MG, Granado-Font E, Ferré-Grau C, Montaña-Carreras X. Mobile phone apps to promote weight loss and increase physical activity: a systematic review and meta-analysis. J Med Internet Res 2015;17(11):e253 [FREE Full text] [doi: 10.2196/jmir.4836]
5. Wu Y, Yao X, Vespasiani G, Nicolucci A, Dong Y, Kwong J, et al. Mobile app-based interventions to support diabetes self-management: a systematic review of randomized controlled trials to identify functions associated with glycemic efficacy. JMIR Mhealth Uhealth 2017 Mar 14;5(3):e35. [doi: 10.2196/mhealth.6522]
6. Anderson K, Emmerton LM. Contribution of mobile health applications to self-management by consumers: review of published evidence. Aust Health Rev 2016 Dec;40(5):591-597. [doi: 10.1071/AH15162] [Medline: 26681206]
7. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791]
8. Kemp A, Jones S, Lawson Z, Maguire S. Patterns of burns and scalds in children. Arch Child 2014 Feb 03;99(4):316-321. [doi: 10.1136/archdischild-2013-304991]
9. Cooper N, Kendrick D, Timblin C, Hayes M, Majsak-Newman G, Meteyard K, et al. The short-term cost of falls, poisonings and scalds occurring at home in children under 5 years old in England: multicentre longitudinal study. Inj Prev 2016 Jan 29;22(5):334-341. [doi: 10.1136/injuryprev-2015-041808]
10. Tse T, Poon C, Tse KH, Tsui TK, Ayyappan T, Burd A. Paediatric burn prevention: an epidemiological approach. Burns 2006 Mar;32(2):229-234. [doi: 10.1016/j.burns.2005.08.025] [Medline: 16448760]
11. Hammig B, Ogletree R. Burn injuries among infants and toddlers in the United States, 1997-2002. Am J Health Behav 2006;30(3):259-267. [doi: 10.5555/ajhb.2006.30.3.259] [Medline: 16712440]
12. Ali K, Spinks J. No tea until three? Inj Prev 2003 Mar;9(1):92-93 [FREE Full text] [Medline: 12642572]
13. Battle C, Evans V, James K, Guy K, Whitley J, Evans P. Epidemiology of burns and scalds in children presenting to the emergency department of a regional burns unit: a 7-year retrospective study. Burns Trauma 2016;4:19 [FREE Full text] [doi: 10.1186/s41038-016-0047-7] [Medline: 27574688]
14. Dewar D, Magson C, Fraser J, Crighton L, Kimble R. Hot beverage scalds in Australian children. J Burn Care Rehabil 2004;25(3):224-227. [Medline: 15273461]
15. Burgess JD, Kimble RM, Cameron CM, Stockton K. Hot beverage scalds in Australian children: still simmering 10 years on. J Burn Care Res 2016;37(4):e335-e339. [doi: 10.1097/BCR.0000000000000267] [Medline: 26132050]
16. Brusselaers N, Monstrey S, Vogelaers D, Hoste E, Blot S. Severe burn injury in Europe: a systematic review of the incidence, etiology, morbidity, and mortality. Crit Care 2010;14(5):R188 [FREE Full text] [doi: 10.1186/cc9300] [Medline: 20958968]
17. Burgess JD, Kimble RM, Watt KA, Cameron CM. Hot tea and tiny tots don't mix: a cross-sectional survey on hot beverage scalds. Burns 2017 Dec;43(8):1809-1816 [FREE Full text] [doi: 10.1016/j.burns.2017.05.008]
18. Rosenstock I. Historical origins of the health belief model. Health Educ Behav 1974 Dec;2(4):328-335. [doi: 10.1177/109019817400200403]
19. Burgess JD, Cameron CM, Watt KA, Kimble RM. Cool Runnings - an app-based intervention for reducing hot drink scalds: study protocol for a randomised controlled trial. Trials 2016 Dec 03;17(1):388 [FREE Full text] [doi: 10.1186/s13063-016-1521-z] [Medline: 27488411]
20. Burgess JD, Kimble RM, Watt KA, Cameron CM. The adoption of social media to recruit participants for the Cool Runnings randomized controlled trial in Australia. JMIR Res Protoc 2017 Oct 24;6(10):e200 [FREE Full text] [doi: 10.2196/resprot.8189] [Medline: 29066427]
21. Hilder L, Zhichao Z, Parker M, Jahan S, Chambers G. Canberra: AIHW; 2014. Australia's mothers and babies 2012. Perinatal statistics series no. 30. Cat. no. PER 69 URL: https://www.aihw.gov.au/getmedia/674fe3d3-4432-4675-8a96-cab97e3c277f/18530.pdf.aspx?inline=true [accessed 2018-09-25] [WebCite Cache ID 72gn9S1na]

22. Hugo Centre for Migration and Population Research. Adelaide. ARIA and Accessibility Research URL: http://www.adelaide.edu.au/apmrc/ [accessed 2017-05-29] [WebCite Cache ID 6rBw4EMIV]
23. ABS. SEIFA 2011 URL: http://www.abs.gov.au/websitedbs/censushome.nsf/home/seifa2011 [accessed 2018-09-21] [WebCite Cache ID 72aZpUdqU]
24. Burgess JD, Watt KA, Kimble RM, Cameron CM. Knowledge of childhood burn risks and burn first aid: Cool Runnings. Inj Prev 2018 Jan 31:1-6 Epub ahead of print(forthcoming). [doi: 10.1136/injuryprev-2017-042650] [Medline: 29386371]
25. Wood F, Phillips M, Jovic T, Cassidy J, Cameron P, Edgar D, Steering Committee of the Burn Registry of AustraliaNew Zealand (BRANZ). Water First Aid Is Beneficial In Humans Post-Burn: Evidence from a Bi-National Cohort Study. PLoS One 2016;11(1):e0147259 [FREE Full text] [doi: 10.1371/journal.pone.0147259] [Medline: 26808839]
26. Triola MM, Triola MJ. Biostatistics for the Biological and Health Sciences with Statdisk. Boston: Pearson; 2013.
27. Turner C, Spinks A, McClure RJ, Nixon J. Community-based interventions for the prevention of burns and scalds in children. Cochrane Database Syst Rev 2004(3):CD004335 [FREE Full text] [doi: 10.1002/14651858.CD004335.pub2] [Medline: 15266531]
28. Scott I. Laws and Rule-making. In: McClure R, Stevenson M, McEvoy S, editors. The Scientific Basis of Injury Prevention and Control. Melbourne: IP Communications; Oct 01, 2004:283-302.
29. Zou K, Wynn P, Miller P, Hindmarch P, Majsak-Newman G, Young B, et al. Preventing childhood scalds within the home: overview of systematic reviews and a systematic review of primary studies. Burns 2015 Aug;41(5):907-924 [FREE Full text] [doi: 10.1016/j.burns.2014.11.002] [Medline: 25841997]
30. DiGuiseppi C, Roberts I. Individual-level injury prevention strategies in the clinical setting. Future Child 2000;10(1):53-82. [Medline: 10911688]
31. Davies M, Maguire S, Okolie C, Watkins W, Kemp AM. How much do parents know about first aid for burns? Burns 2013 Sep;39(6):1083-1090. [doi: 10.1016/j.burns.2012.12.015] [Medline: 23347890]
32. Kimble R, Griffin B. Reducing the incidence of burn injuries to Indigenous Australian children. Med J Aust 2017 May 15;206(9):389-390. [doi: 10.5694/mja17.00151]
33. Poulos R, Hayen A, Finch C, Zwi A. Area socioeconomic status and childhood injury morbidity in New South Wales, Australia. Inj Prev 2007 Oct;13(5):322-327 [FREE Full text] [doi: 10.1136/ip.2007.015693] [Medline: 17916889]
34. Edelman L. Social and economic factors associated with the risk of burn injury. Burns 2007 Dec;33(8):958-965. [doi: 10.1016/j.burns.2007.05.002] [Medline: 17869003]
35. King L, Thomas M, Gatenby K, Georgiou A, Hua M. "First aid for scalds" campaign: reaching Sydney's Chinese, Vietnamese, and Arabic speaking communities. Inj Prev 1999 Jun;5(2):104-108 [FREE Full text] [Medline: 10385828]
36. Cagle K, Davis J, Dominic W, Gonzales W. Results of a focused scald-prevention program. J Burn Care Res 2006;27(6):859-863. [doi: 10.1097/01.BCR.0000245423.79531.50] [Medline: 17091083]
37. Poushter J. Pew Global. Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies URL: http://www.pewglobal.org/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/ [accessed 2018-09-21] [WebCite Cache ID 72akm3vJU]
38. Queensland Government. 2015. Queensland Perinatal Data Collection URL: https://www.health.qld.gov.au/__data/assets/pdf_file/0029/631847/Peri_intro2015.pdf [accessed 2017-08-09] [WebCite Cache ID 6sgk0G3r0]
39. Monica A. Pew Internet. The Demographics of Device Ownership URL: http://www.pewinternet.org/2015/10/29/the-demographics-of-device-ownership/ [accessed 2017-07-10] [WebCite Cache ID 6xoJuyFF8]
40. Ferguson B. Games for Wellness-Impacting the Lives of Employees and the Profits of Employers. Games Health J 2012 Jun;1(3):177-179. [doi: 10.1089/g4h.2012.0023] [Medline: 26193434]
41. GamesIndustry. 8 billion in 2016 URL: http://www.gamesindustry.biz/articles/2012-05-21-gamification-market-to-reach-usd2-8-billion-in-2016 [accessed 2017-09-14] [WebCite Cache ID 71ntXCuAh]
42. Lister C, West JH, Cannon B, Sax T, Brodegard D. Just a fad? Gamification in health and fitness apps. JMIR Serious Games 2014 Aug 04;2(2):e9 [FREE Full text] [doi: 10.2196/games.3413] [Medline: 25654660]
43. Larsson L. Stockholm, Sweden: Kth Royal Institute of Technology; 2015. Framework for Increased Intrinsic Motivation using Gamification Strategies in E-Learning Internet URL: http://www.diva-portal.se/smash/get/diva2:861526/FULLTEXT01.pdf [accessed 2018-08-20] [WebCite Cache ID 71nuz8JQs]

Abbreviations

ARIA: Accessibility/Remoteness Index in Australia
IQR: interquartile range
NNT: number needed to treat
OR: odds ratio
RCT: randomized controlled trial
SEIFA: Socioeconomic Indexes for Areas
SES: socioeconomic status

Edited by G Eysenbach; submitted 10.03.18; peer-reviewed by E Omaki, A Samaranayaka; comments to author 29.03.18; revised version received 28.05.18; accepted 21.08.18; published 10.10.18.

Please cite as:
Burgess J, Watt K, Kimble RM, Cameron CM
Combining Technology and Research to Prevent Scald Injuries (the Cool Runnings Intervention): Randomized Controlled Trial
J Med Internet Res 2018;20(10):e10361
URL: http://www.jmir.org/2018/10/e10361/
doi:10.2196/10361
PMID:30305263

Original Paper

Promoting Psychological Well-Being at Work by Reducing Stress and Improving Sleep: Mixed-Methods Analysis

Denny Meyer[1,2*], BSc (Hons), MBL, DBL; Madawa W Jayawardana[1*], BSc (Hons), PhD; Samuel D Muir[2*], BSc (Hons), MSc; David Yen-Teh Ho[3*], BSc; Olivia Sackett[1*], BEnvSc (Hons), PhD

[1]Department of Statistics, Data Science and Epidemiology, Swinburne University of Technology, Melbourne, Australia

[2]Centre of Mental Health, Faculty of Health, Arts and Design, Swinburne University of Technology, Melbourne, Australia

[3]Virgin Pulse, Boston, MA, United States

[*]all authors contributed equally

Corresponding Author:
Denny Meyer, BSc (Hons), MBL, DBL
Department of Statistics, Data Science and Epidemiology
Swinburne University of Technology
John Street
Hawthorn
Melbourne, VIC 3122
Australia
Phone: 61 392144824
Fax: 61 392145260
Email: dmeyer@swin.edu.au

Abstract

Background: Workplace programs designed to improve the health and psychological well-being of employees are becoming increasingly popular. However, there are mixed reports regarding the effectiveness of such programs and little analysis of what helps people to engage with such programs.

Objective: This evaluation of a particularly broad, team-based, digital health and well-being program uses mixed methods to identify the elements of the program that reduce work stress and promote psychological well-being, sleep quality, and productivity of employees.

Methods: Participation in the Virgin Pulse Global Challenge program during May to September 2016 was studied. Self-reported stress, sleep quality, productivity, and psychological well-being data were collected both pre- and postprogram. Participant experience data were collected through a third final survey. However, the response rates for the last 2 surveys were only 48% and 10%, respectively. A random forest was used to estimate the probability of the completion of the last 2 surveys based on the preprogram assessment data and the demographic data for the entire sample (N=178,350). The inverse of these estimated probabilities were used as weights in hierarchical linear models in an attempt to address any estimation bias caused by the low response rates. These linear models described changes in psychological well-being, stress, sleep, and productivity over the duration of the program in relation to gender and age, engagement with each of the modules, each of the program features, and participant descriptions of the Virgin Pulse Global Challenge. A 0.1% significance level was used due to the large sample size for the final survey (N=18,653).

Results: The final analysis suggested that the program is more beneficial for older people, with 2.9% greater psychological well-being improvements observed on average in the case of women than men ($P<.001$). With one exception, all the program modules contributed significantly to the outcome measures with the following average improvements observed: psychological well-being, 4.1%-6.0%; quality of sleep, 3.2%-6.9%; work-related stress, 1.7%-6.8%; and productivity, 1.9%-4.2%. However, only 4 of the program features were found to have significant associations with the outcome measures with the following average improvements observed: psychological well-being, 3.7%-5.6%; quality of sleep, 3.4%-6.5%; work-related stress, 4.1%-6.4%; and productivity, 1.6%-3.2%. Finally, descriptions of the Virgin Pulse Global Challenge produced 5 text topics that were related to the outcome measures. Healthy lifestyle descriptions showed a positive association with outcomes, whereas physical activity and step count tracking descriptions showed a negative association with outcomes.

Conclusions: The complementary use of qualitative and quantitative survey data in a mixed-methods analysis provided rich information that will inform the development of this and other programs designed to improve employee health. However, the low response rates and the lack of a control group are limitations, despite the attempts to address these problems in the analysis.

(J Med Internet Res 2018;20(10):e267) doi:10.2196/jmir.9058

KEYWORDS

exercise; productivity; healthy lifestyle

Introduction

Background

Workplace health initiatives promoting behavioral change have been recognized as vitally important strategies for improving the health of employees [1-3]. A workplace health and exercise program can be broadly defined as an intervention implemented in a workplace environment, targeting employees to improve health behaviors such as increasing physical activity, eating better, and taking care of their mental health [4]. However, a meta-analytic review has found that the effectiveness of such programs differs, with universal programs being more effective than targeted programs, one-on-one delivery being more effective than group-based classroom delivery, and with Web-based delivery and train-the-trainer delivery producing nonsignificant improvements [5]. Comparisons between Web- and paper-based workplace programs have shown similar retention rates and improvements; however, a randomized control trial (RCT) has shown higher popularity ratings for Web-based programs than for paper-based programs [6].

In an earlier meta-analytic review of workplace health promotion programs [7], 18 studies describing 21 RCT interventions found little overall effect for workplace health promotion programs (d=0.24, 95% CI 0.14-0.34). The effectiveness was larger in younger populations, in interventions with weekly contacts, and in studies where the control group received no health promotion. This systematic review highlights the importance of sound methodologies for statistical analyses and the handling of missing data in studies of this nature [7]. It was found that when an intention-to-treat analysis was performed, a 2.6-fold lower effectiveness was observed with a 1.7-fold lower effectiveness for studies controlling for confounders. In addition, studies of poor methodological quality reported a 2.9-fold higher effect size for workplace health promotion programs. Workplace health and exercise as well as workplace health promotion studies often tend to be observational in nature, without the luxury of a randomly assigned control group. This means that there is no way to determine if the results are a consequence of taking the program or due to the confounding factors. Moreover, attrition rates tend to be very high, making the likelihood of biased estimation results very high. In particular, more engaged participants are likely to be over-represented in the final results, providing an overoptimistic assessment of program efficacy. However, modern missing value methods have started addressing this problem [8-10]. In particular, the use of inverse probability weights (IPWs) is recommended when data are missing for large numbers of variables [8], and mixed model or hierarchical (multilevel) linear models are recommended over multiple imputation methods when repeated outcome measures are missing.

However, studies have shown that missing data in mHealth are closely linked to the level of engagement [11], suggesting that the inclusion of engagement measures in outcome studies may also help to address the issue of estimation bias caused by high levels of attrition.

Objectives

Much of the research has focused on the overall effectiveness of workplace health and exercise programs and workplace health promotion programs rather than investigating the characteristics of more successful programs and investigating the characteristics of employees for whom such programs are more or less beneficial. It is this gap in the literature that this research attempts to fill, using a workplace health and exercise program entitled the Virgin Pulse Global Challenge (VPGC). In this study, we use modern missing value approaches to model outcomes of interest, including various measures of engagement as predictors in these models.

The evaluation of interventions usually involves a quantitative comparison of baseline and postassessment and/or follow-up performance using measures relevant to the intervention focus. However, postassessment surveys usually collect qualitative as well as quantitative data by way of open-ended questions. Although these qualitative data may be reported in a descriptive sense, there is seldom any attempt to incorporate this information into the analysis of how and why an intervention may fail or succeed. Due to the recent availability of sophisticated text mining tools, it is now possible to augment quantitative evaluations of an intervention with qualitative data using *mixed methods* [12,13]. In this paper, we use this approach with participant descriptions of the VPGC, demonstrating that mixed-methods evaluations can result in richer information for both researchers and commercial organizations alike.

In this paper, we address the following 4 hypotheses:

> H1: Demographic effects will be associated with improvements in psychological well-being, sleep quality, stress, and productivity.
>
> H2: Program module effects will be associated with improvements in psychological well-being, sleep quality, stress, and productivity.
>
> H3: Program feature effects will be associated with improvements in psychological well-being, sleep quality, stress, and productivity.
>
> H4: Program descriptions will be associated with improvements in psychological well-being, sleep quality, stress, and productivity.

Methods

Virgin Pulse Global Challenge

Virgin Pulse is a global Software-as-a-Service vendor providing several health and well-being programs designed to improve the psychological well-being of employees. The Global Challenge is one of the Virgin Pulse programs that features a team-based health and well-being challenge. The challenge consists of a 100-day virtual journey around the world, referred to as the 100-Day Journey. As part of the 100-Day Journey, employees are placed in teams of 7 individuals from their organization and provided with an activity tracker (the Pulse Device or other third-party-supported devices) and access to an app that is available through Web browsers and on mobile devices. Teams compete with one another to accumulate steps measured by their activity trackers.

The VPGC program differs from most other workplace health and exercise programs in terms of its breadth. In particular, the simultaneous focus on social, physical, and mental health is regarded as a strength of this program, which is seldom seen in other programs. In addition to promoting physical activity, the program incorporates a number of modules that focus on encouraging improvement in sleep, nutrition, and psychological well-being. The Balance module addresses mental health issues, and the Heart Age module provides 2 evaluations: a lifestyle score out of 1000 and heart age relative to real age.

The program is gamified to encourage employees to develop healthy habits through education, goal setting, and positive reinforcement using progress monitoring and achievement awards (eg, virtual trophies).

Participants

The target population for this research was participants from all the organizations that were enrolled in the VPGC program that commenced in May 2016. Participants agreed to the use of their personal data by any agencies engaged by Virgin Pulse for the purposes of quality control. They did this when they signed up for the VPGC program on the internet [14], when encouraged to do so by their employers. The nature of this encouragement differed for each employer and is therefore not reported here. No Virgin Pulse incentives were offered to employees to participate.

The VPGC platform and its practices around data security and privacy have been externally audited and certified against the following standards: ISO 27001:2013, TRUSTe privacy seal, and General Data Protection Regulation governing data protection and privacy. The data were deidentified and password protected before being made available to the researchers and were held on university password-protected computers. Ethics approval for the evaluation of this program by the Swinburne University of Technology was obtained from the Swinburne University Research Ethics Committee (SHR Project 2017/061).

Surveys

The data were automatically collected using 3 electronic Web-based surveys, administered on the internet as closed (password-protected) voluntary surveys. The initial Web-based survey was completed early in May 2016, the second Web-based survey was completed toward the end of the 100-Day Journey, and the final Participant Experience Survey (PES) was completed 2 weeks later. There was no randomization of items in any of these Web-based surveys and no adaptive questioning. The first Web-based survey included 28 questions and the second survey included 29 questions, all with a Likert scale (0-6) or Yes/No responses. For the PES, there were 27 questions in a variety of formats (text, multiple, and single response answers). Questions were presented to users in approximate groupings of *pages* to drive an easy and engaging completion process. All questions were voluntary, and there was no review step.

Measures

To measure psychological well-being, this study used the independently validated World Health Organization 5-item questionnaire (WHO-5) on psychological well-being. A total of 5 simple and noninvasive questions constitute this measure of subjective psychological well-being, which has been validated as a sensitive and specific screening tool for depression. This scale was first published in 1998; it has been translated into 30 languages and used all over the world [15]. Responses to these items were used to calculate an overall score, where 0 is the “worst imaginable” and 100 the “best imaginable” psychological well-being.

In this study, the WHO-5 score was used as the primary outcome measure. Secondary outcome measures were self-assessed levels of work-related stress, sleep quality, and productivity for the last month (measured on a 0-6 ordinal scale pre- and postprogram). In all cases, higher scores indicated a more desirable state. All the above outcome measures were collected at the start (T1) and end (T2) of the VPGC program.

In the final PES (T3), an attempt was made to identify the engagement factors perceived to be particularly beneficial by the participants. In particular, participant perceptions were considered with regard to (1) the best program modules (ie, Physical Activity, Heart Age, Sleep, Nutrition, and Balance) and (2) the best program features (eg, virtual trophies, the leaderboard, individual and team challenges, and connection with colleagues). These variables were all measured on a binary scale (0 for a negative response and 1 for a positive response). In this final T3 survey, participants also provided a response to the following question: “How would you describe the Global Challenge to a friend or colleague?” As described below, these responses were used to create 25 text topic scores for each respondent, consisting of values between 0 and 1 [16].

Response Rates

Response rates often have little meaning in the context of workplace health and exercise programs and workplace health promotion programs [17] because surveys may only be partially completed. In Figure 1, the number of responses to the WHO-5 questions, representing our primary outcome measure, are summarized for the first (T1) and second (T2) surveys. A completion rate of 85% is suggested for the first survey and 48% for the second survey. The final PES (T3) was completed by only 10.5% of all participants. This low response rate suggests that there will be an estimation bias in any models fitted using the T3 data unless modern methods addressing this bias are applied. Even then, results must be viewed with caution.

Statistical Analysis

The analysis was divided into 4 phases. In phase 1, descriptive statistics were presented for the final (T3) PES. Phase 2 involved predicting completion of this T3 survey and the WHO-5 psychological well-being measure for the second (T2) survey, using data collected in the first (T1) survey (including any missing data for the first survey). This predicted probability was inverted to produce the IPWs that were used in the ensuing analyses to reduce any estimation bias arising from missing data. Phase 3 consisted of the text mining used to produce the 25 topics and topic scores relating to the final PES question: “How would you describe the Global Challenge to a friend or colleague?” In the fourth phase, hierarchical linear model analyses were conducted for each of the outcome measures using IPW to address the problem of missing data. These analyses allow the research hypotheses to be addressed while attempting to adjust for any estimation bias caused by the low response rates. Phase 1 was conducted with IBM SPSS Version 25 software. Phase 2 and phase 3 analyses were conducted using the SAS Institute Enterprise Miner Version 14.2, whereas R software was used to produce a word cloud for the responses to the VPGC descriptions. Phase 4 was conducted using SSI Central HLM7 software.

Phase 1: Descriptive Statistics for Participant Experience Survey Responses

Responses for males and females were compared using independent samples *t* tests for the outcome measures and Fisher exact tests for the crosstab tests.

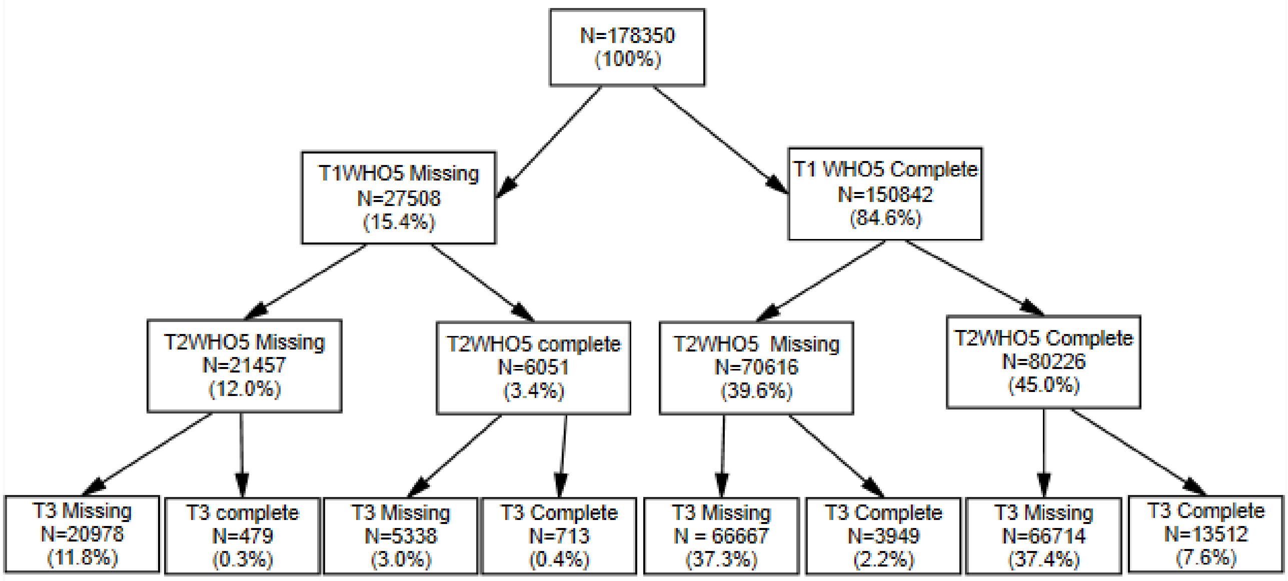
Phase 2: Estimated Inverse Probability Weights

Various tools were considered for modeling the completion of the PES (T3) and the T2 WHO-5 in terms of the first survey (T1) responses. In particular, a random forest was compared with single trees, gradient boosting, a neural network with 3 hidden nodes, support vector machines, and probit/logistic regression analyses [18-20]. All 178,350 of the original participants were randomly split into training, validation, and test datasets in a 40:30:30 ratio. To compensate for the low completion rate, the models were optimized in terms of profit, with a one unit profit for each correctly identified missing survey response and a 10 unit profit for each correctly identified completed survey response. This was done using the training data. The classification accuracy of the various tools was compared with the validation and test data to ensure reliable reproducible results. For the goodness-of-fit measure, the area under the receiver operating characteristic (ROC) curve was used for the test data, with higher values indicative of a better fit. The random forest consisted of 100 trees each constructed using a random 60% of the data, with chi-square tests used to choose the optimum splitting variable for each split. This method ensured that the random forests were not overfitted and that nonlinear relationships were accommodated. Missing values were treated as distinct categories in this analysis.

Phase 3: Text Mining

Text mining was applied to analyze unstructured responses to the question “How would you describe the Global Challenge to a friend or colleague?”, producing text topics [16,21]. The first step involved the extraction of terms from the text, followed by an automatic filtering of terms that were too frequent or too uncommon. A spectral decomposition was then conducted to determine which terms commonly occurred together in the same response. The most important 25 topics were extracted, and scores were then assigned to each participant for each of these topics, with values ranging from 0 to 1. Higher values indicate a more likely topic description of the VPGC.

Figure 1. Survey participation (T1=first survey, T2=second survey, and T3=final participant experience survey). WHO-5: World Health Organization 5-item questionnaire.



Phase 4: Hierarchical Linear Model Analyses

The hypotheses were then addressed for each of the outcome measures using a hierarchical (multilevel) linear model analysis. For all these analyses, the IPWs calculated in phase 2 were applied to correct the estimation bias caused by low response rates. These hierarchical linear models [22] used all available data for the T3 sample, including cases with missing data, and they allowed for correlation between measures collected in the first 2 surveys using a maximum likelihood approach. In view of the large sample size (N=18,653), only significance levels of less than .001 are considered to be significant in the tables provided below.

Results

Phase 1: Descriptive Statistics for Participant Experience Survey (T3) Respondents

Table 1 shows that for the T3 respondents, the Physical Activity module was most popular, with more than 90% of participants, whereas the Heart Age module was popular with more than 60% of participants. Females appreciated the Balance module more than males, although the size of this effect was very small (φ<0.1).

Table 1. Gender comparison for Participant Experience Survey (T3) respondents.

Responses	Female (N=10,397)	Male (N=8256)	Total (N=18,653)	*P* value	Effect size
Age in years, mean (SD)	42.51 (10.97)	43.46 (10.44)	42.93 (10.75)	<.001	*d*=0.09
Helpful modules, n (%)					
Physical Activity	9393 (90.33)	7474 (90.54)	16,867 (90.41)	.67	φ=0.00
Heart Age	6380 (61.36)	4905 (59.41)	11,285 (60.50)	.007	φ=0.02
Sleep	2314 (22.24)	1911 (23.15)	4225 (22.64)	.15	φ=.01
Nutrition	3153 (30.33)	2481 (30.04)	5634 (30.20)	.68	φ=0.00
Balance	3941 (37.90)	2765 (33.47)	6706 (35.94)	<.001	φ=0.05
Helpful features, n (%)					
Mini-challenge	6731 (64.73)	4771 (57.79)	11,502 (61.66)	<.001	φ=0.07
Leaderboard	5110 (49.12)	4429 (53.67)	9539 (51.13)	<.001	φ=0.05
Competitions	2747 (26.41)	2471 (29.96)	5218 (27.98)	<.001	φ=0.04
My_Location	4515 (43.43)	2855 (34.56)	7370 (39.51)	<.001	φ=0.09
Individual mini-leagues	1875 (18.02)	1652 (20.02)	3527 (18.90)	<.001	φ=0.03
Team mini-leagues	2316 (22.25)	1882 (22.82)	4198 (22.50)	.41	φ=0.01
Trophies	4429 (42.59)	3194 (38.70)	7623 (40.87)	<.001	φ=0.04
Communication sharing	738 (7.10)	480 (5.81)	1218 (6.53)	<.001	φ=0.03
My_Stats	5951 (57.20)	4989 (60.45)	10,940 (58.64)	<.001	φ=0.03
More colleague connect	8125 (78.90)	6512 (79.44)	14,637 (79.14)	.40	φ=0.01
Outcomes T1, mean (SD)					
WHO-5[a] (0-100)	52.86 (19.02)	56.06 (18.97)	54.27 (19.06)	<.001	*d*=0.17
Quality of sleep (0-6)	3.20 (1.21)	3.25 (1.18)	3.22 (1.19)	.004	*d*=0.04
Reduced work stress (0-6)	2.99 (1.38)	3.08 (1.31)	3.03 (1.35)	<.001	*d*=0.07
Productivity (0-6)	3.85 (1.01)	3.85 (1.00)	3.85 (1.01)	.78	*d*=0.00
Outcomes T2, mean (SD)					
WHO-5 (0-100)	67.70 (18.48)	69.15 (18.37)	68.33 (18.45)	<.001	*d*=0.08
Quality of sleep (0-6)	3.99 (1.14)	4.01 (1.14)	4.00 (1.14)	.18	*d*=0.02
Reduced work stress (0-6)	3.76 (1.43)	3.80 (1.38)	3.78 (1.41)	.11	*d*=0.03
Productivity (0-6)	4.24 (1.01)	4.21 (1.02)	4.23 (1.01)	.04	*d*=0.03

[a] WHO-5: World Health Organization 5-item questionnaire.

Table 2. Pearson correlations for outcome measures at T1 and T2, with T2 correlations italicized in the Lower Triangular Matrix.

Outcomes	Psychological well-being	Quality of sleep	Reduced work stress	Productivity
Psychological well-being	—	0.452[a]	0.420[a]	0.419[a]
Quality of sleep	*0.571* [a]	—	0.246[a]	0.207[a]
Reduced work stress	*0.533* [a]	*0.373* [a]	—	0.168[a]
Productivity	*0.54* [a]	*0.392* [a]	*0.342* [a]	—

[a] $P<.001$.

As shown in Table 1, improved connections because of the program were claimed by 79% of the participants, suggesting that the 7-person team feature of the VPGC is particularly effective. However, there was a significant but surprisingly small association between improved connections with colleagues and endorsement of the Physical Activity module ($P<.001$, $\varphi=.07$), suggesting that shared physical activity was probably not contributing very much to this improvement in colleague connections. Other features such as the mini-challenges, the leaderboard used for comparing the performance of all teams, and personal daily step count performance (My_Stats) were also important. There were several significant but very small gender differences in the case of program feature preferences, with females favoring the mini-challenges, My_Location (virtual travel related to step count performance), and Trophy features more than males, whereas males tended to favor the competitive program features more than women.

As shown in Table 2, there were significant correlations between the outcome measures at T1 and T2, with correlations of moderate size with psychological well-being and weaker correlations between quality of sleep, reduced work-related stress, and productivity.

As shown in Table 1, on average, males scored higher than females in terms of psychological well-being (WHO-5). Males scored significantly better than females in terms of sleep quality and work-related stress only at the time of the first survey (T1). However, these effect sizes were again very small ($d<0.2$). The above descriptive statistics suggest improvements in the 4 outcome measures over the duration of the program, with large to moderate effect sizes (η^2) of 0.37, 0.30, 0.23, and 0.11 for psychological well-being, sleep, work-related stress, and productivity, respectively. However, the low response rates for the T2 and particularly the T3 sample make any such claims premature. To address this issue of nonresponse bias, IPWs were calculated as indicated below, and a hierarchical (multilevel) regression analysis was conducted in phase 4.

Phase 2: Estimated Inverse Probability Weights

Perhaps not surprisingly, the random forest produced the best results for predicting completed responses, with an area under the ROC curve of 0.719 for the test data. However, single trees were not far behind, with areas under the ROC curve of 0.704. Other methods (support vector machines, gradient boosting, neural networks, and binary regression) produced disappointing results, with areas of less than 0.62 in all cases.

Figure 2 illustrates how probability predictions are obtained for a single (Gini) tree with splits occurring in such a way as to minimize the heterogeneity in any node in relation to survey completion. This heterogeneity is measured using the Gini criterion [20]. The “Count” in this figure refers to the number of participants for the training and validation data in each node, with the thick black line indicating the path for the majority of participants after each split. A code of 1 indicates a missing survey response, whereas a 0 indicates a completed survey response.

In the random forest and the single tree shown in Figure 2, missing values for the *weight* (kg) variables were particularly useful for identifying missing survey responses. It seems that people who refused to supply their weight in the first survey (T1) are unlikely to complete 1 or more of the last 2 surveys. Younger age (<50 years) or missing age was also highly predictive of missing survey responses. The third most important variable for the random tree was psychological well-being (WHO-5 T1). A missing or very low WHO-5 T1 score (<22) was also associated with failure to complete surveys, suggesting that people who are depressed are more likely to drop out. This result confirms that the estimation bias is inevitable if analyses are conducted on the completer data, without making any attempt to account for this bias. The inverse of the estimated completion probabilities obtained from the random forest model was therefore used as weights (IPWs) to control for bias in the hierarchical (multilevel) regression models described below.

Phase 3: Text Mining

The PES included the following question: “In one or two sentences, how would you describe the Global Challenge to a friend or colleague?” Multimedia Appendix 1 shows a word cloud for the responses to this question, whereas parsing, filtering, and topic selection produced the 25 topics displayed in Figure 3. The number of terms and the number of respondents (#Docs) associated with each topic are also displayed in Figure 3, with the highest number of respondents (1002) owning topic 11, relating to “good motivation” and “good fun.”

Phase 4: Hierarchical Linear Model Analyses

Table 3 provides the results for independent hierarchical linear model analyses addressing each of the hypotheses. Only significant program features and text topics are included in Table 3. In terms of demographic effects, the program appears to be more beneficial for older people, and in the case of psychological well-being, it seems that women benefit more than men, with average improvements being 2.9% higher for women than men.

Figure 2. Single tree for predicting the probability for survey completion. WHO-5: World Health Organization 5-item questionnaire.

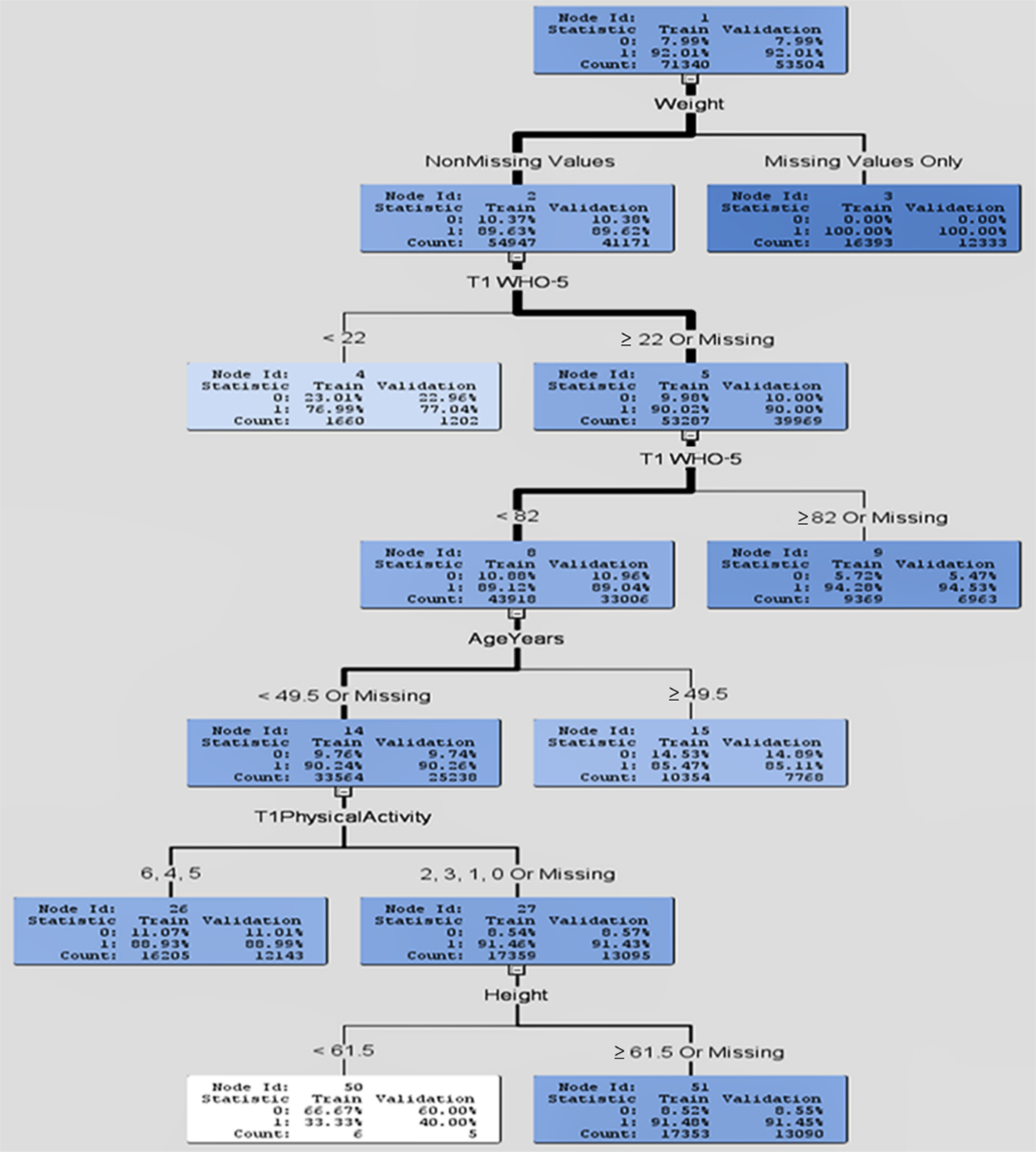
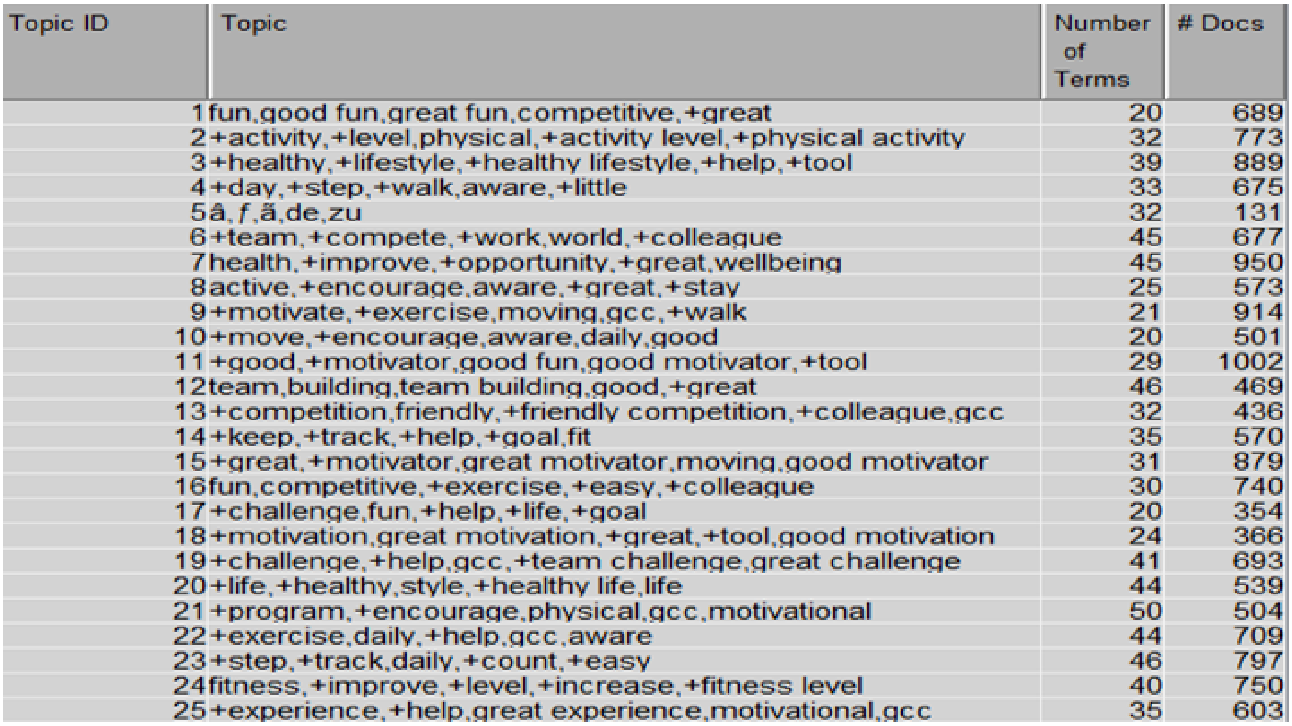


Figure 3. Topics extracted from the Global Challenge description.



Topic ID	Topic	Number of Terms	# Docs
1	fun,good fun,great fun,competitive,+great	20	689
2	+activity,+level,physical,+activity level,+physical activity	32	773
3	+healthy,+lifestyle,+healthy lifestyle,+help,+tool	39	889
4	+day,+step,+walk,aware,+little	33	675
5	â,ƒ,ã,de,zu	32	131
6	+team,+compete,+work,world,+colleague	45	677
7	health,+improve,+opportunity,+great,wellbeing	45	950
8	active,+encourage,aware,+great,+stay	25	573
9	+motivate,+exercise,moving,gcc,+walk	21	914
10	+move,+encourage,aware,daily,good	20	501
11	+good,+motivator,good fun,good motivator,+tool	29	1002
12	team,building,team building,good,+great	46	469
13	+competition,friendly,+friendly competition,+colleague,gcc	32	436
14	+keep,+track,+help,+goal,fit	35	570
15	+great,+motivator,great motivator,moving,good motivator	31	879
16	fun,competitive,+exercise,+easy,+colleague	30	740
17	+challenge,fun,+help,+life,+goal	20	354
18	+motivation,great motivation,+great,+tool,good motivation	24	366
19	+challenge,+help,gcc,+team challenge,great challenge	41	693
20	+life,+healthy,style,+healthy life,life	44	539
21	+program,+encourage,physical,gcc,motivational	50	504
22	+exercise,daily,+help,gcc,aware	44	709
23	+step,+track,daily,+count,+easy	46	797
24	fitness,+improve,+level,+increase,+fitness level	40	750
25	+experience,+help,great experience,motivational,gcc	35	603

The program modules have significant relationships with improvements in outcomes for psychological well-being (4.1%-6.0% on average), quality of sleep (3.2%-6.9% on average), work-related stress (1.7%-6.8% on average), and productivity (1.9%-4.2% on average). However, engagement with the Physical Activity module is not significantly related to improvements in work-related stress. The Nutrition program has the strongest association with improvements in psychological well-being, the Sleep module has the strongest association with improvements in the quality of sleep, and the Balance module has the strongest association with improvements in work-related stress and productivity.

However, as shown in Table 3, only 4 of the program features are related to improvements in the outcome measures: for psychological well-being (3.7%-5.6% on average), quality of sleep (3.4%-6.5% on average), work-related stress (4.1%-6.4% on average), and productivity (1.6%-3.2% on average). Improvements in connection with colleagues, the mini-challenges, and (virtual) trophies are all associated with significant improvements in all 4 outcome measures, whereas sharing (Web-based talk) with the VPGC community is associated with significant improvements in the quality of sleep and work-related stress. There are also only 5 text topics that appear to be significantly associated with changes in the outcome measures: for psychological well-being (−3.9% to 5.2% on average), quality of sleep (−4.4% to 5.0% on average), work-related stress (−5.9% to 7.0% on average), and productivity (−1.4% to 2.7% on average).

Text topic 2 relates to physical levels of activity, and the negative association with quality of sleep suggests that participants with this perception of the program saw a decline in their quality of sleep over the duration of the program. Text topic 3 relates to a healthy lifestyle, and topic 15 relates to the perception of the VPGC as a good or great motivator. The results suggest that for people with these perceptions of the program, there was evidence of an improvement in psychological well-being and quality of sleep. Topic 20 again relates to a healthy lifestyle, suggesting that stress at work is also reduced for these people. Finally, topic 23 relates to the daily tracking of step counts, and this perception of the program is associated with negative changes in psychological well-being and stress at work.

Table 4 compares the various hierarchical linear models in terms of variance explained. The time effect refers to differences between the first 2 surveys. Clearly, this time effect explains a significant amount of the variance with R^2 values ranging from 3.3% in the case of productivity to 11.9% in the case of psychological well-being. However, adding demographic variables, engagement with program modules, and features and perceptions of the program, as described to a friend or colleague, makes little difference to these R^2 values, with an increase of only 2.5% for psychological well-being, 0.7% for quality of sleep, 1.4% for stress at work, and 2.2% for productivity. This means that although some support has been found for all 4 hypotheses, the size of these effects is small.

Table 3. Estimated program effects.

Improvements in outcome measures	Psychological well-being	Quality of sleep	Work-related stress	Productivity
T1 outcomes, range	0-100	0-6	0-6	0-6
T1, mean (SD)	54.27 (19.06)	3.22 (1.19)	3.03 (1.35)	3.85 (1.01)
Demographic effects, estimated coefficients (% change with respect to T1)				
Age in years	0.069 (0.1)[a]	0.007 (0.2)[a]	0.004 (0.1)[a]	0.001 (0.0)
Female	1.595 (2.9)[a]	0.022 (0.7)	0.055 (1.8)	0.031 (0.8)
Module effects, estimated coefficients (% change with respect to T1)				
Physical activity	3.17 (5.8)[a]	0.140 (4.3)[a]	0.052 (1.7)	0.137 (3.6)[a]
Heart age	2.24 (4.1)[a]	0.103 (3.2)[a]	0.099 (3.3)[a]	0.075 (1.9)[a]
Sleep	3.10 (5.7)[a]	0.263 (8.2)[a]	0.157 (5.2)[a]	0.131 (3.4)[a]
Nutrition	3.26 (6.0)[a]	0.154 (4.8)[a]	0.132 (4.4)[a]	0.131 (3.4)[a]
Balance	3.21 (5.9)[a]	0.222 (6.9)[a]	0.205 (6.8)[a]	0.160 (4.2)[a]
Feature effects, estimated coefficients (% change with respect to T1)				
Connections	2.965 (5.5)[a]	0.172 (5.3)[a]	0.124 (4.1)[a]	0.114 (3.0)[a]
Mini-challenge	3.012 (5.6)[a]	0.108 (3.4)[a]	0.131 (4.3)[a]	0.125 (3.2)[a]
Trophies	2.479 (4.6)[a]	0.133 (4.1)[a]	0.146 (4.8)[a]	0.071 (1.8)[a]
Community sharing	2.02 (3.7)	0.209 (6.5)[a]	0.195 (6.4)[a]	0.062 (1.6)
Text topics, estimated coefficients (% change with respect to T1)				
#2	−0.158 (0.3)	−0.142 (4.4)[a]	−0.026 (0.9)	0.031 (0.8)
#3	1.938 (3.6)[a]	0.127 (3.9)[a]	0.063 (2.1)	0.079 (2.1)
#15	2.224 (4.1)[a]	0.110 (3.4)[a]	0.024 (0.8)	0.076 (2.0)
#20	2.806 (5.2)[a]	0.160 (5.0)[a]	0.213 (7.0)[a]	0.103 (2.7)
#23	−2.116 (3.9)[a]	−0.062 (1.9)	−0.180 (5.9)[a]	−0.054 (1.4)

[a] $P<.001$.

Table 4. Proportion of variance explained.

Predictors	R^2 (%)			
	Psychological well-being	Quality of sleep	Work-related stress	Productivity
Time effect (T1-T2)	11.9	9.7	7.1	3.3
Time effect with demographics	12.8	9.8	7.4	3.9
Time effect with modules	12.6	10.2	7.6	4.1
Time effect with features	13.5	10.4	7.9	4.4
Time effect with text topics	12.6	10.0	7.4	4.3
All variables	14.6	10.4	8.5	5.5

[a] Psychological well-being, quality of sleep, reduced work stress, productivity.

Discussion

Overview

This study has identified program modules and features of a workplace health and exercise program that are particularly helpful to employees, with differences observed between men and women, and more benefit for older people. Although these gender effects are small, they are significant. Two methods have been used in an attempt to address the low response rates for the second and third surveys. A random forest has been used to create IPWs, and these weights have been utilized in hierarchical (multilevel) models, utilizing maximum likelihood methods to minimize the estimation bias resulting from missing data. The study has used text mining to incorporate qualitative data in the hierarchical linear models, which is rarely seen [23].

Principal Findings

The research hypotheses are all supported to some extent as explained below.

1. The VPGC program is associated with greater improvements in psychological well-being, quality of sleep, and work-related stress in the case of older employees. Greater improvements in psychological well-being were found for female employees than for male employees.
2. All the modules contributed positively to psychological well-being, quality of sleep, work-related stress, and productivity, with 1 exception. The association was not significant for the Physical Activity module in the case of work-related stress.
3. Employee perceptions for 3 of the program features were significantly associated with improvements for all 4 outcome measures. These 3 features were connections with colleagues nurtured using team structures, the mini-challenges, and (virtual) trophies. VPGC community sharing (Web-based talk) was associated with improvements in quality of sleep and improved levels of work-related stress.
4. Descriptions of the program by participants provide additional context. In particular, it was found that physical activity levels had a negative association with quality of sleep, whereas daily tracking of step counts had a negative association with psychological well-being and stress at work. However, perceptions of the program as a great motivator for a healthy lifestyle were associated with improvements in psychological well-being and stress at work.

The combination of methods used in this analysis provides a better understanding of how the VPGC program may achieve behavioral change. The results suggest that although the Physical Activity module of the program is the most popular, it does not make a significant contribution to reduced work-related stress, and perhaps through its emphasis on step tracking, it has a negative association with sleep and psychological well-being as well. However, there were many positive associations, which suggest that the other modules and several of the program features are associated with positive outcomes. Perceptions of the VPGC as a tool for motivating a healthy lifestyle are especially conducive to positive outcomes.

It has been recommended that modules addressing nutrition and mental health are particularly advantageous and that a variety of program features are beneficial to address the differing preferences of men and women. The VPGC program appears to be more effective with older participants, and future work is required to explain this. However, no rigorous evaluation of the effectiveness of the VPGC is possible on account of the data limitations presented below.

Limitations

Survey completion rates were particularly low for the final (T3) survey, which is crucial for this analysis because it contains the engagement data with the program modules and features and the description of the VPGC data used to create the text topics. With so many variables missing for 90% of the T3 data, IPW was the only way to address the threat of estimation bias. Furthermore, hierarchical linear models were needed to address the problem of missing T2 outcome values for many of the T3 respondents, ensuring that all T3 participants could be retained for the analysis. However, it is still not certain that estimation bias has been avoided. The predictors in our model (eg, helpfulness of modules and features) relate to engagement only indirectly. Other studies involving Web-based programs have used more direct engagement measures, such as the number of sessions completed [24], time to last engagement with the website, or the number of hits or time spent online [25], whereas physical activity programs have used step counts completed [26]. Future studies experiencing low response rates for final surveys should attempt to incorporate these direct measures of engagement as control variables.

Moreover, as there is no control group in this study, it is not possible to claim that the program and its individual modules are beneficial because we have no participants outside of the VPGC. In addition, none of the effects considered in this analysis were strong and must therefore be treated with caution in view of the limitations described above.

Finally, there were no follow-up data that could be used to assess the long-term effects of the program. Future studies should allow for a control group, ideally utilizing an RCT and should provide follow-up data to address these limitations.

Comparison With Prior Work

The VPGC program is an internet-based program that previous research suggests is not the best way to conduct a workplace health and exercise program [5]. However, the results of this study indicate that for those who did complete the PES (T3), the VPGC program is a success. We have been unable to pinpoint exactly why this is the case, but it is thought that the breadth of the program is a contributing factor.

Age was identified as a significant predictor of engagement with the VPGC. Specifically, the results of this study showed that in terms of stress, sleep, and psychological well-being, the VPGC is more successful with older people. This is consistent with some previous workplace health and exercise program evaluations, which have reported that older employees tend to remain more engaged in workplace health and exercise programs in comparison with younger employees [27-30]. However, it is less consistent with other reports [7].

Results of this study confirmed that improving connections with colleagues is a particularly important feature of the VPGC. Previous mixed-methods and qualitative studies have reported that team-based workplace health and exercise programs increase motivation for exercise due to not wanting to let the team down and creating positive topics of conversation among employees [31,32]. This suggests that future workplace health and exercise programs should consider incorporating more social features for participants to engage in, in an attempt to foster greater social support among participants. Previous research has found consistent, positive associations between social support and physical activity [33]. There was a significant association between improvements in social connections and engagement with the Physical Activity module in this study, but this effect was small (φ=0.07).

Although the Physical Activity module was the most popular module, engagement with this module had no significant association with work-related stress. In addition, descriptions of the VPGC relating to physical activity had a negative association with quality of sleep, and the results suggested that the step-tracking component of the Physical Activity module might, for some participants, detract from psychological well-being. However, further investigation is required to determine why this is the case.

The Balance module, which aimed to promote psychological well-being, was found to be an important module with regard to reduced stress and enhanced sleep quality, productivity, and psychological well-being, with the Nutrition module also strongly associated with psychological well-being. This suggests that future workplace health and exercise programs may benefit from incorporating modules focusing on mental health and nutrition, rather than just targeting physical activity.

Conclusions

According to our results for the participants who completed the final T3 survey, the VPGC program is associated with reduced work-related stress, improved quality of sleep, and improved productivity. It is also associated with increases in psychological well-being, especially in the case of women. The qualitative analysis identified a healthy lifestyle as a beneficial perception of the program, whereas the quantitative analysis indicated that the Nutrition and Balance modules contribute the most to program outcomes. However, despite the Physical Activity module being the most popular module, its contribution to reduced work-related stress appears to be limited. The social and gamified features of the program, especially the mini-challenges, appear to make the program a lot of fun.

However, these results must all be regarded as preliminary because of the lack of a control group, the low response rate for the final PES, and the lack of follow-up measures. Further work is required to provide greater certainty.

Acknowledgments

Virgin Pulse is to pay for any publication costs, but no other financial agreements are in place between the authors or their respective employers.

Authors' Contributions

DYTH sourced the data, DM conducted the statistical analysis, SDM produced the literature review, and all authors helped with the writing and checking of this manuscript.

Conflicts of Interest

DYTH and OS were employees of Virgin Pulse when this paper was first submitted. They have not been involved in the analysis. Their contribution has been in providing the data and in ensuring that the program description is correct in this manuscript, thereby avoiding any conflict of interest. There is no conflict of interest for DM, MWJ, or SDM.

Multimedia Appendix 1

Word cloud for descriptions of the Global Challenge.

[PNG File, 177KB - jmir_v20i10e267_fig3.png]

References

1. World Health Organization. 2010. Global recommendations on physical activity for health URL: http://www.who.int/dietphysicalactivity/physical-activity-recommendations-18-64years.pdf [WebCite Cache ID 72ONCsi8k]
2. Dishman RK, Oldenburg B, O'Neal H, Shephard RJ. Worksite physical activity interventions. Am J Prev Med 1998 Nov;15(4):344-361. [Medline: 9838977]
3. Avey JB, Luthans F, Jensen SM. Psychological capital: a positive resource for combating employee stress and turnover. Hum Res Manage 2009 Sep;48(5):677-693. [doi: 10.1002/hrm.20294]
4. Gebhardt DL, Crump C. Employee fitness and wellness programs in the workplace. Am Psychol 1990 Feb;45(2):262-272. [Medline: 2178506]
5. Vanhove AJ, Herian MN, Perez AL, Harms PD, Lester PB. Can resilience be developed at work? A meta-analytic review of resilience-building programme effectiveness. J Occup Organ Psychol 2016;89:278-307 [FREE Full text] [doi: 10.1111/joop.12123]
6. Cook RF, Billings DW, Hersch RK, Back AS, Hendrickson A. A field test of a web-based workplace health promotion program to improve dietary practices, reduce stress, and increase physical activity: randomized controlled trial. J Med Internet Res 2007 Jun 19;9(2):e17 [FREE Full text] [doi: 10.2196/jmir.9.2.e17] [Medline: 17581811]

7. Rongen A, Robroek SJ, van Lenthe FJ, Burdorf A. Workplace health promotion: a meta-analysis of effectiveness. Am J Prev Med 2013 Apr;44(4):406-415. [doi: 10.1016/j.amepre.2012.12.007] [Medline: 23498108]
8. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. Stat Methods Med Res 2013 Jun;22(3):278-295. [doi: 10.1177/0962280210395740] [Medline: 21220355]
9. Siddiqui O. MMRM versus MI in dealing with missing data—a comparison based on 25 NDA data sets. J Biopharm Stat 2011 May;21(3):423-436. [doi: 10.1080/10543401003777995] [Medline: 21442517]
10. Peters SA, Bots ML, den Ruijter HM, Palmer MK, Grobbee DE, Crouse 3rd JR, METEOR study group. Multiple imputation of missing repeated outcome measurements did not add to linear mixed-effects models. J Clin Epidemiol 2012 Jun;65(6):686-695. [doi: 10.1016/j.jclinepi.2011.11.012] [Medline: 22459429]
11. Scherer EA, Ben-Zeev D, Li Z, Kane JM. Analyzing mHealth engagement: joint models for intensively collected user engagement data. JMIR Mhealth Uhealth 2017 Jan 12;5(1):e1 [FREE Full text] [doi: 10.2196/mhealth.6474] [Medline: 28082257]
12. Creswell JW. A Concise Introduction to Mixed Methods Research (Sage Mixed Methods Research). Los Angeles: SAGE Publications, Inc; 2014.
13. Johnson RB, Onwuegbuzie AJ, Turner LA. Toward a definition of mixed methods research. J Mix Methods Res 2007;1(2):112-133. [doi: 10.1177/1558689806298224]
14. Virgin Pulse Global Challenge. URL: https://globalchallenge.virginpulse.com/ [WebCite Cache ID 6tjrr0YoW]
15. Topp CW, Østergaard SD, Søndergaard S, Bech P. The WHO-5 Well-Being Index: a systematic review of the literature. Psychother Psychosom 2015;84(3):167-176 [FREE Full text] [doi: 10.1159/000376585] [Medline: 25831962]
16. Chakraborty G, Pagolu M, Garla S. Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS. Cary, North Carolina, USA: SAS Institute; 2014.
17. 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]
18. Berry MJ, Linoff GS. Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management. New York: Wiley, John & Sons, Incorporated; 2011.
19. Shannon CE. A mathematical theory of communication. Bell Syst Tech J 1948;27(3):379-423 [FREE Full text] [doi: 10.1002/j.1538-7305.1948.tb01338]
20. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction (Springer Series in Statistics). New York, NY: Springer-Verlag; 2009.
21. Dumais ST. Latent semantic analysis. Ann Rev Info Sci Tech 2005 Sep 22;38(1):188-230. [doi: 10.1002/aris.1440380105]
22. Raudenbush SW. In: Bryk AS, editor. Hierarchical Linear Models: Applications and Data Analysis Methods, Vol. 1. Thousand Oaks, CA: SAGE Publications; 2001.
23. Palinkas LA. Qualitative and mixed methods in mental health services and implementation research. J Clin Child Adolesc Psychol 2014;43(6):851-861 [FREE Full text] [doi: 10.1080/15374416.2014.910791] [Medline: 25350675]
24. Usher-Smith JA, Winther LR, Shefer GS, Silarova B, Payne RA, Griffin SJ. Factors associated with engagement with a web-based lifestyle intervention following provision of coronary heart disease risk: mixed methods study. J Med Internet Res 2017 Oct 16;19(10):e351 [FREE Full text] [doi: 10.2196/jmir.7697] [Medline: 29038095]
25. Morrison C, Doherty G. Analyzing engagement in a web-based intervention platform through visualizing log-data. J Med Internet Res 2014 Nov 13;16(11):e252 [FREE Full text] [doi: 10.2196/jmir.3575] [Medline: 25406097]
26. Davies C, Corry K, Van Itallie A, Vandelanotte C, Caperchione C, Mummery WK. Prospective associations between intervention components and website engagement in a publicly available physical activity website: the case of 10,000 Steps Australia. J Med Internet Res 2012 Jan 11;14(1):e4 [FREE Full text] [doi: 10.2196/jmir.1792] [Medline: 22260810]
27. Abraham JM, Feldman R, Nyman JA, Barleen N. What factors influence participation in an exercise-focused, employer-based wellness program? Inquiry 2011;48(3):221-241. [doi: 10.5034/inquiryjrnl_48.03.01] [Medline: 22235547]
28. Abraham JM, Crespin DJ, Rothman AJ. Initiation and maintenance of fitness center utilization in an incentive-based employer wellness program. J Occup Environ Med 2015 Sep;57(9):952-957 [FREE Full text] [doi: 10.1097/JOM.0000000000000498] [Medline: 26340283]
29. Mattila E, Orsama AL, Ahtinen A, Hopsu L, Leino T, Korhonen I. Personal health technologies in employee health promotion: usage activity, usefulness, and health-related outcomes in a 1-year randomized controlled trial. JMIR Mhealth Uhealth 2013 Jul 29;1(2):e16 [FREE Full text] [doi: 10.2196/mhealth.2557] [Medline: 25098385]
30. Robroek SJ, Lindeboom DE, Burdorf A. Initial and sustained participation in an internet-delivered long-term worksite health promotion program on physical activity and nutrition. J Med Internet Res 2012 Mar 05;14(2):e43 [FREE Full text] [doi: 10.2196/jmir.1788] [Medline: 22390886]
31. Edmunds S, Stephenson D, Clow A. The effects of a physical activity intervention on employees in small and medium enterprises: a mixed methods study. Work 2013 Jan 01;46(1):39-49. [doi: 10.3233/WOR-121523] [Medline: 23241703]
32. Scherrer P, Sheridan L, Sibson R, Ryan MM, Henley N. Employee engagement with a corporate physical activity program: the global corporate challenge. J Int Bus Stud 2010;18(1):125-139 [FREE Full text]

33. Trost SG, Owen N, Bauman AE, Sallis JF, Brown W. Correlates of adults' participation in physical activity: review and update. Med Sci Sports Exerc 2002 Dec;34(12):1996-2001. [doi: 10.1249/01.MSS.0000038974.76900.92] [Medline: 12471307]

Abbreviations

IPWs: inverse probability weights
PES: Participant Experience Survey
RCT: randomized controlled trial
ROC: receiver operating characteristic
VPGC: Virgin Pulse Global Challenge
WHO: World Health Organization

Edited by G Eysenbach; submitted 26.09.17; peer-reviewed by A Burdorf, P Romero-Sanchiz; comments to author 18.12.17; revised version received 04.06.18; accepted 21.06.18; published 19.10.18.

Please cite as:
Meyer D, Jayawardana MW, Muir SD, Ho DYT, Sackett O
Promoting Psychological Well-Being at Work by Reducing Stress and Improving Sleep: Mixed-Methods Analysis
J Med Internet Res 2018;20(10):e267
URL: https://www.jmir.org/2018/10/e267/
doi:10.2196/jmir.9058
PMID:30341045

Original Paper

A Mobile Videoconference-Based Intervention on Stress Reduction and Resilience Enhancement in Employees: Randomized Controlled Trial

Johanna Inyang Kim[1], MD, PhD; Je-Yeon Yun[2,3], MD, PhD; Heyeon Park[4], PhD; Suk-Young Park[4], MA; Youngsheen Ahn[5], MA; Hansol Lee[6], MA; Tae-Kwon Kim[7], MS, MBA; Sooran Yoon[7], MA; Young-Joon Lee[7], MD, PhD; Sohee Oh[8], PhD; John W Denninger[9,10], MD, PhD; Bung-Nyun Kim[2,3,11], MD, PhD; Jeong-Hyun Kim[4], MD, PhD

[1]Hanyang University Medical Center, Seoul, Republic Of Korea
[2]Seoul National University Hospital, Seoul, Republic Of Korea
[3]Seoul National University College of Medicine, Seoul, Republic Of Korea
[4]Seoul National University Bundang Hospital, Seongnam, Republic Of Korea
[5]Seoultop Psychiatric Clinic, Seoul, Republic Of Korea
[6]Korean Academy of Gifted Education, Seoul, Republic Of Korea
[7]Aimmed Company Ltd, Seoul, Republic Of Korea
[8]Seoul Metropolitan Government Seoul National University Boramae Medical Center, Seoul, Republic Of Korea
[9]Benson-Henry Institue for Mind Body Medicine, Massachusetts General Hospital, Boston, MA, United States
[10]Department of Psychiatry, Harvard Medical School, Boston, MA, United States
[11]Human Behavioral Medical Institute, Seoul National University Medical Research Center, Seoul, Republic Of Korea

Corresponding Author:
Jeong-Hyun Kim, MD, PhD
Seoul National University Bundang Hospital
82, Gumi-ro 173 beon-gil, Bundang-gu, Seongnam-si, Gyeonggi-do 13620
Seongnam,
Republic Of Korea
Phone: 82 31 787 2025
Email: retrial3@hanmail.net

Abstract

Background: Videoconferencing-based treatments have shown great potential in increasing engagement and compliance by decreasing the barriers of time and distance. In general, employees tend to experience a lot of stress, but find it difficult to visit a clinic during office hours.

Objective: The purpose of this study was to investigate the effectiveness of a mobile videoconference-based intervention for stress reduction and resilience enhancement in employees.

Methods: In total, 81 participants were randomly allocated to one of the three conditions: mobile videoconferencing, in-person, and self-care; of these, 72 completed the study. All participants underwent assessment via self-reported questionnaires before, immediately after, and 1 month after the intervention. Intervention lasted for 4 weeks and consisted of elements of cognitive behavioral therapy, positive psychology, and meditation. Changes in clinical variables regarding stress and resilience across time were compared between treatment conditions.

Results: There were significant condition × time effects on variables measuring perceived stress, resilience, emotional labor, and sleep, demonstrating significantly differential effects across time according to treatment condition. Moreover, there were significant effects of condition on perceived stress and occupational stress. There were no significant differences in any variable between the mobile videoconferencing and in-person conditions at 1 month after the intervention.

Conclusions: Results indicate that both mobile videoconferencing and in-person interventions were comparably effective in decreasing stress and enhancing resilience. Further studies with a larger sample size and a longer follow-up period are warranted to investigate the long-term effect of mobile videoconferencing interventions.

Trial Registration: ClinicalTrials.gov identifier NCT03256682; https://clinicaltrials.gov/ct2/show/NCT03256682 (Archived by WebCite at http://www.webcitation.org/71W77bwnR)

(J Med Internet Res 2018;20(10):e10760) doi:10.2196/10760

KEYWORDS

employees; mobile phone; randomized controlled trial; resilience enhancement; stress reduction; videoconferencing

Introduction

Stress is a major public health concern. It can cause serious psychological and physical problems such as fatigue, sleeping problems, coronary diseases, depression, and even related mortality [1-4]. Hence, proper stress management for employees is particularly important; work-related stress affects approximately 28% of all workers [5]. Stress has also been reported to be a major factor in up to 80% of all work-related injuries and 40% of workplace turnovers [6]. Data from the American Institute of Stress show that job stress in the United States accounts for over US $300 billion annually as a result of increased absenteeism, employee turnover, diminished productivity, medical, legal, and insurance expenses, and worker compensation [7].

Traditional face-to-face or in-person stress reduction interventions for employees are proven to be effective, with small-to-moderate effect sizes [8]. However, a major barrier to managing stress is limited access due to time and location constraints, as it may be difficult for employees to find the time to see a therapist during work hours on weekdays [9]. The success of nonpsychopharmacological psychiatric treatment is related to treatment adherence, which is linked to barriers like time and distance. A useful way to increase access, and thereby increase treatment adherence, is by incorporating telemedicine-based methods. Employees can easily access videoconferencing-based treatments at their convenience, ie, during break hours in the workplace or after office hours at home [10].

There is a bulk of evidence suggesting that videoconferencing-based telepsychology is no less effective than in-person treatment in a variety of psychiatric disorders [11], including depression [12], panic disorders [13], obsessive compulsive disorder [14], posttraumatic stress disorder [15], and eating disorders [16]. A recent meta-analysis of 26 randomized controlled trials demonstrated noninferiority of remote psychiatric counseling, with respect to both assessment and treatment, compared with in-person counseling [17]. To the best of our knowledge, to date, there has been no study evaluating the effectiveness of videoconferencing in reducing stress and enhancing resilience in the workplace.

An important issue in videoconferencing is equipment. In most previous studies, participants were instructed to use the equipment provided at remote videoconference sites [11]. However, traveling to these sites is no different from traveling to see a counselor; it would still be a barrier to treatment for those whose workplace or home is far from the videoconference site. To further increase accessibility to treatment, we used mobile-based videoconferencing equipment. Access to treatment by smartphones could be particularly beneficial in South Korea, since almost all Korean employees own a smartphone (94% of population), making it one of the countries with the highest smartphone ownership worldwide [18].

The purpose of this study was to compare the effects of mobile videoconference-based intervention on stress reduction and resilience enhancement with that of in-person and self-care methods in Korean employees. We hypothesized that the effectiveness of the mobile videoconference condition would be superior to the control condition (self-care condition), but comparable to the in-person condition.

Methods

Participants

Participants were recruited via advertisements at the Seoul National University Hospital and the Seoul National University Bundang Hospital between August 2017 and November 2017 (NCT03256682). The inclusion criteria were (1) age between 19 and 65 years, (2) a score of ≥14 on the Perceived Stress Scale (PSS) at baseline, (3) possession of an Android smartphone, and (4) currently employed full-time. As stress in the workplace is commonly associated with depression, insomnia, and anxiety [1,19] and as stress management interventions have been found to be effective at treating these three conditions [20,21], we included individuals with these conditions as long as the type and dosage of their medication had not changed for the past 6 months. The exclusion criteria were (1) age <19 or >65 years; 2) cognitive disorders, such as intellectual disability or dementia; 3) neurological disorders, including epilepsy, stroke, or others; 4) history of schizophrenia or bipolar I disorder; 5) current report of suicidal ideation; and 6) nonpharmacological treatment or counseling within the past 6 months.

During the screening process, psychiatric diagnoses were confirmed using the Mini-International Neuropsychiatric Interview (MINI), which is a short, structured psychiatric interview designed to detect a wide range of Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition and International Classification of Diseases, Tenth Revision psychiatric disorders [22]. Questions are phrased to allow only “yes” or “no” answers. The Korean version has well-established validity and reliability [23]. MINI was conducted by two psychologists with a master’s degree in education and 2 hours of training in advance.

With a predicted effect size of Cohen *d*=0.4, an alpha level of.05, a desired power of 0.95, and a correlation of 0.5 among repeated measures, the estimated total sample size using G-Power was 69 (23 participants per condition). Considering a drop-out rate of 20%, we aimed to recruit 87 participants.

Written informed consent was obtained from all participants after sufficient explanation of the study. The study protocol was approved by the Institutional Review Board of Seoul National University Bundang Hospital.

Assessments

Demographic information, including age, gender, length of work (<3 years or ≥3 years), marital status (married and not separated, others), and education status (college education or more, less than a college education), was obtained using a self-reported questionnaire. These demographic variables have been previously found to relate to occupational stress level [24]. The questionnaires were filled on paper. The primary outcome measures of this study were changes in scores of PSS and Brief Resilience Scale (BRS) scales. Changes in scores of scales assessing emotional labor, occupational stress, and insomnia were evaluated as secondary outcomes.

PSS is a 10-item questionnaire used to assess perceived stress [25]. It was designed to measure the degree to which respondents found their life situations unpredictable, uncontrollable, and overbearing. It includes 10 direct queries about incidents that brought upon states of being upset, nervous, stressed, or irritated; four items were worded negatively and the rest positively. Answers were given on a 5-point Likert scale (0: never to 4: very often). Total scores were calculated after reversing the scores from positive items and then summing up all scores. Higher scores indicated higher levels of perceived stress. The Korean version of PSS was found to have a Cronbach alpha=.74 and a test-retest reliability of intraclass correlation coefficient=0.81 [26].

We used the Korean Emotional Labor Scale (KELS) [27] to measure the level of emotional labor. Emotional labor is defined as the process by which workers have to control their feelings in accordance with the organizational demand and occupational role [28,29]. KELS was developed to measure Korean-specific emotional labor and was validated with a nation-wide random sample of 1042 Korean employees. It was based on the literature related to emotional labor [29-35]; emotional labor scales such as the Emotional Labor Inventory [36], Emotional Labor Scale [37], Emotion Work Requirements Scale [38], and Frankfurt Emotion Work Scale [39]; and a focused group interview. KELS has 5 subscales: effort to control emotion (5 items), organizational monitoring system (4 items), demands of emotional labor (3 items), emotional damage (6 items), and organizational support system (7 items). Each item was rated on a 4-point Likert scale, from 1 (not at all) to 4 (very much). Scores for each subscale were calculated based on the scoring method provided by the developers. The possible range for each subscale was 0-100, with higher scores representing higher levels of emotional labor. In this study, we only used the total score.

The level of job stress was measured using the Korean Occupational Stress Scale (KOSS)-Short Form, which is one of the most commonly used questionnaires for assessing job stress in South Korea [40]. It consists of 24 items measured on a 4-point Likert scale (1: never to 4: always). This scale comprises 7 subscales, including job demand, job control, interpersonal conflict, job insecurity, organizational system, lack of reward, and workplace environment. The sum of each subscale was calculated and then converted to 100 points. We used the KOSS total score in the analysis; higher scores indicated a higher level of job stress.

BRS was used to measure individual resilience [41]. It aims to assess the most traditional and original sense of resilience, in other words, “the ability to bounce back from stress [42].” It consists of 6 items measured on a 5-point Likert scale (1: strongly disagree to 5: strongly agree). While other resilience scales measure personal characteristics that may promote positive adaptation, BRS is the only scale that targets and assesses resilience itself.

Insomnia was measured by the Athens Insomnia Scale (AIS), which contains 8 items scored on a 4-point Likert scale [43]. Total AIS scores range from 0 to 24, with higher scores indicating greater symptom severity. AIS has been validated for screening insomnia in South Korean firefighters with good psychometric properties (Cronbach alpha=.88 and item-total correlation=0.73) [44].

All self-reported questionnaires were completed by the participants before treatment, immediately after treatment, and 1 month after treatment.

At posttreatment, participants answered 4 questions about therapeutic alliance. Each item was rated on a 5-point Likert scale (1: disagree strongly to 5: agree strongly). The 4 questions were as follows: (1) “I felt as if the therapist understood me well”; (2) “I felt as if the therapist was paying attention to what I was saying”; (3) “I could tell that the therapist was empathetic by his/her tone of voice”; and (4) “I felt comfortable during therapy sessions.”

Randomization and Treatment Conditions

We performed 1:1:1 block randomization (stratified by organization) with randomly selected block sizes (3, 6, or 9) using REDCap (Research Electronic Data Capture) tools hosted at the Medical Research Collaborating Center of Seoul National University Bundang Hospital. REDCap generates randomization codes using SAS software. The allocation sequence was concealed to the patients until they had entered the trial and to the investigators until the end of the study. Participants in the mobile videoconference and in-person conditions underwent 50-minute sessions of 1:1 therapy with 1 of 3 psychologists with a master’s degree in education for 4 weeks (one session a week). The protocol of therapy was adapted from the Stress Management and Resilience Training: Relaxation Response Resilience Program (SMART-3RP) [45]. SMART-3RP is an 8-week, 1.5-hour session program developed by the Benson-Henry Institute for Mind Body Medicine at Massachusetts General Hospital. This program is based on the principles of cognitive behavioral therapy and positive psychology in conjunction with methods that elicit a relaxation response. The goals of the program include (1) eliciting a relaxation response through meditation, (2) reducing overall stress reactivity, and (3) increasing connectedness to oneself and others. In this study, we modified the SMART-3RP program into a 4-week program (1-hour per session); a brief summary of each session is presented in Multimedia Appendix 1. The participants in the self-care condition received educational material regarding methods to self-regulate stress and were instructed to read 1 chapter each week for 4 weeks. This material was also provided to the participants in other conditions.

Apparatus

For mobile videoconferencing we used the “Hello Mindcare” Android app [46], which was developed to provide mobile counseling services. All participants downloaded the app free of charge. Using Web Real-Time Communication, the Hello Mindcare app provides videoconferencing by allowing direct peer-to-peer communication and eliminating the need to install plugins or ActiveX. With a highly secure system, all data shared during videoconference sessions were encoded using Transport Layer Security, 128-bit block encryption algorithm ARIA (Academy, Research Institute, Agency), and Advanced Encryption Standard. The Hello Mindcare app provides a booking system, videoconferencing, document sharing, and workbooks for clients to fill in directly via their smartphones. Therapists used the Hello Mindcare counselor Web to manage schedules, participate in videoconferencing sessions, and check participants’ workbooks.

Statistical Analysis

The demographic and clinical characteristics of each condition were compared using analysis of variance (ANOVA) for continuous variables and chi-square tests or Fisher exact tests for categorical variables. Condition, time, and condition × time effects on clinical variables were tested using repeated measure ANOVA with age and marital status included as covariates. For post hoc analyses, we conducted pairwise comparisons of the changes in scores at posttreatment (posttreatment score − pretreatment score) and at 1-month follow-up (follow-up score − pretreatment score) by analysis of covariance, with age and marital status included as covariates. All statistical analyses were performed using IBM SPSS Statistics version 22.0 software (IBM Corp, Chicago, IL, USA). A two-tailed *P* value of <.05 was considered statistically significant.

Results

Of the 98 individuals screened for this study, 17 did not meet the inclusion criteria; 81 individuals were thus enrolled and randomly allocated to 1 of the 3 conditions. Among them, 4 in the mobile videoconference condition and 1 in the in-person condition dropped out after randomization but before treatment initiation (mobile videoconference condition: 3 had trouble installing the app on their smartphone and 1 refused participation due to difficulty in scheduling appointments; in-person condition: 1 needed psychiatric treatment due to aggravation of psychiatric symptoms). As a result, 21 participants were allocated to the mobile videoconference condition, 27 to the in-person condition, and 28 to the self-care condition; all 81 subjects completed the pretreatment assessment. After the start of treatment, 3 participants in the mobile videoconference condition and 1 participant in the self-care condition dropped out (mobile videoconferencing condition: 2 dropped out due to their personal schedules, 1 complained of unstable Wi-Fi connection; self-care condition: 1 dropped out of because of personal matters, but refused to give a detailed explanation). A total of 18 individuals in the videoconference condition, 27 in the in-person condition, and 27 in the self-care condition completed all 4 sessions of the intervention and underwent the posttreatment and 1-month follow-up assessment (Figure 1). The drop-out rates after treatment engagement were 14% (3/21), 0%, and 3% (1/27) for the mobile videoconferencing, in-person, and self-care conditions, respectively; this distribution was not statistically significant (*P*=.09 by Fisher exact test).

The demographic and clinical characteristics of participants who completed the assessments at all three time points are presented in Table 1. The mean age of participants in the mobile videoconference condition and in-person condition was higher than that of the self-care condition (*P*<.001), and there was a significant difference in marital status among the conditions (*P*=0.03); hence, age and marital status were included as covariates in the main analyses. There were no significant differences in gender, length of work, education status, or baseline scores of the clinical variables. When classifying the occupations of the participants according to the International Standard Classification of Occupations 08, 5 were managers, 26 were professionals, 6 were technicians and associate professionals, 12 were clerical support workers, 19 were service and sales workers, 1 was a plant and machine operator, and 3 had elementary occupations. Overall, 42% (30/72) were hospital employees.

The effects of condition, time, and condition × time for all clinical variables are shown in Multimedia Appendix 2; Figures 2-5 depict changes in the PSS, KELS, BRS, and AIS scores across time. The interaction between time and condition was significant for 4 clinical variables (PSS: $F_{3.40}$=3.1, *P*=.03; BRS: $F_{3.47}$=3.9, *P*=.008; KELS: $F_{3.45}$=2.8, *P*=.03; AIS: $F_{3.04}$=4.5, *P*=.005). There were significant main effects for condition on PSS (F_2=8.7, *P*<.001) and KOS (F_2=11.6, *P*<.001).

At posttreatment, the mobile videoconferencing condition showed a greater decrease in KELS scores at posttreatment, but this was not significant at follow-up (Multimedia Appendix 2). There were no significant differences in any clinical variable between the mobile videoconferencing condition and in-person condition at follow-up. The mobile videoconferencing condition showed a greater decrease in KELS scores compared with the self-care condition at posttreatment, but this was without significance at follow-up. The mobile videoconferencing condition showed a greater increase in BRS scores at follow-up than the self-care condition. The in-person condition showed a greater decrease in PSS, KOSS, BRS, and AIS scores at posttreatment than the self-care condition, but only KOSS, BRS, and AIS scores were significant at follow-up.

Regarding the 4 questions about therapeutic alliance, there were no differences in the scores between the videoconferencing condition and in-person condition: question 1: average score 4.7 (SD 0.6) versus 4.7 (SD 0.4), *P*=.58; question 2: average score 4.7 (SD 0.6) versus 4.9 (SD 0.3), *P*=.32; question 3: average score 4.6 (SD 0.5) versus 4.9 (SD 0.3), *P*=.06; question 4: average score 4.7 (SD 0.6) vs 4.7 (SD 0.6), *P*=.98.

Figure 1. Flowchart of the study process. f/u: follow-up.

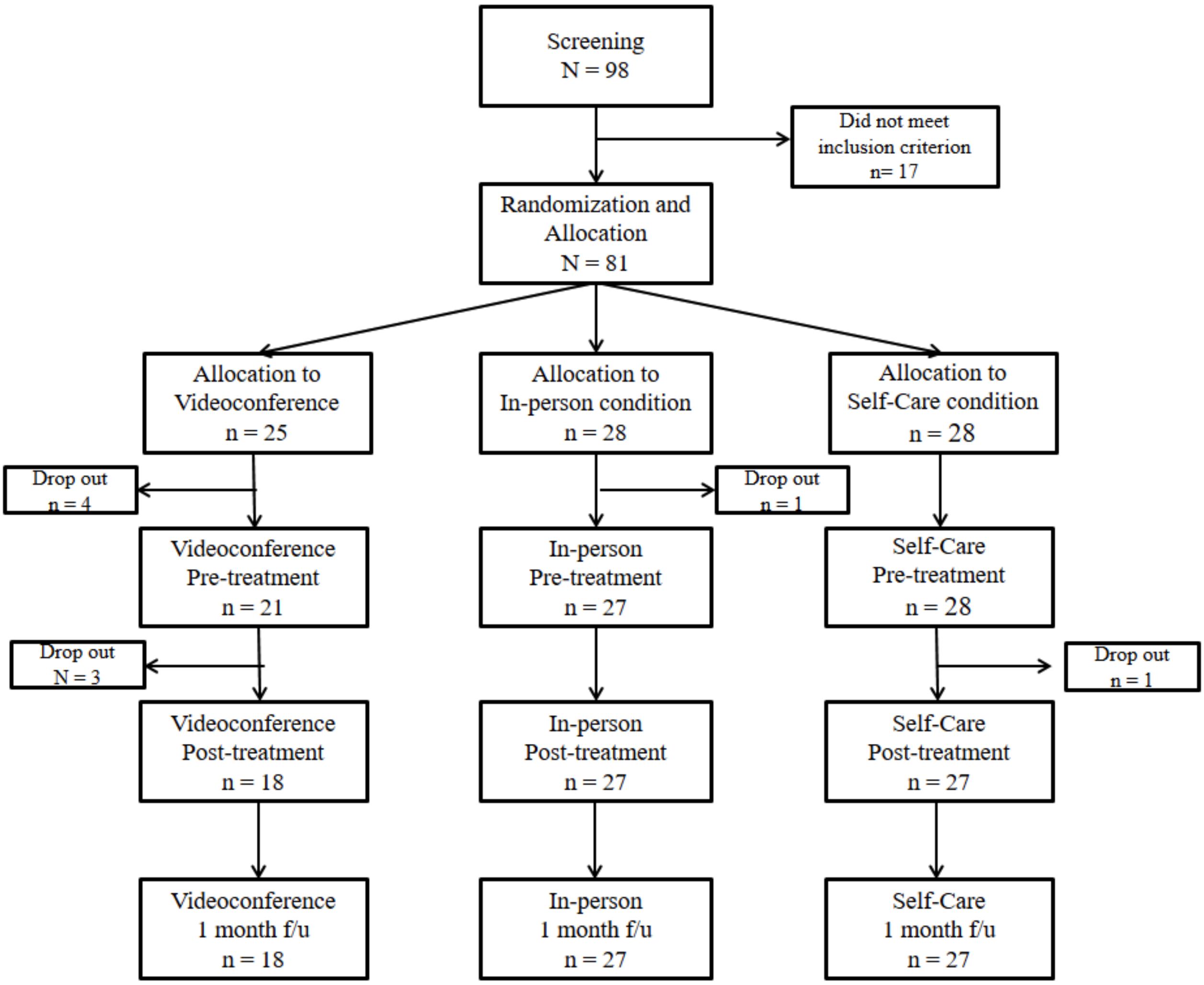


Table 1. Demographic and clinical characteristics of participants in videoconferencing, in-person, and self-care conditions.

Characteristic	Videoconference (n=18)	In-person (n=27)	Self-care (n=27)	*P* value
Age in years, mean (SD)	36.2 (9.2)	36.7 (10.3)	46.6 (9.6)	<.001[a]
Sex (female), n (%)	17 (94.4)	25 (92.6)	25 (92.6)	.97
Education ≥college bachelor degree, n (%)	16 (88.9)	24 (88.9)	18 (66.7)	.07
Marital status (married and unseparated), n (%)	10 (55.6)	17 (63.0)	24 (88.9)	.03
Employed >3 years, n (%)	14 (77.8)	14 (51.9)	19 (70.4)	.16
Perceived Stress Scale, mean (SD)	23.5 (4.2)	23.0 (3.3)	24.6 (4.1)	.30
Brief Resilience Scale, mean (SD)	16.3 (4.1)	16.9 (3.5)	17.2 (4.0)	.75
Korean Emotional Labor Scale, mean (SD)	61.1 (17.1)	57.1 (15.6)	59.1 (14.6)	.70
Korean Occupational Stress Scale, mean (SD)	52.8 (7.7)	53.1 (9.5)	57.5 (9.3)	.12
Athens Insomnia Scale, mean (SD)	16.3 (3.2)	16.7 (4.2)	16.6 (4.1)	.94

[a] Post hoc test by least significant difference; age was less in videoconference and in-person conditions than in self-care conditions.

Figure 2. Change in Perceived Stress Scale scores (PESS) across time according to condition.

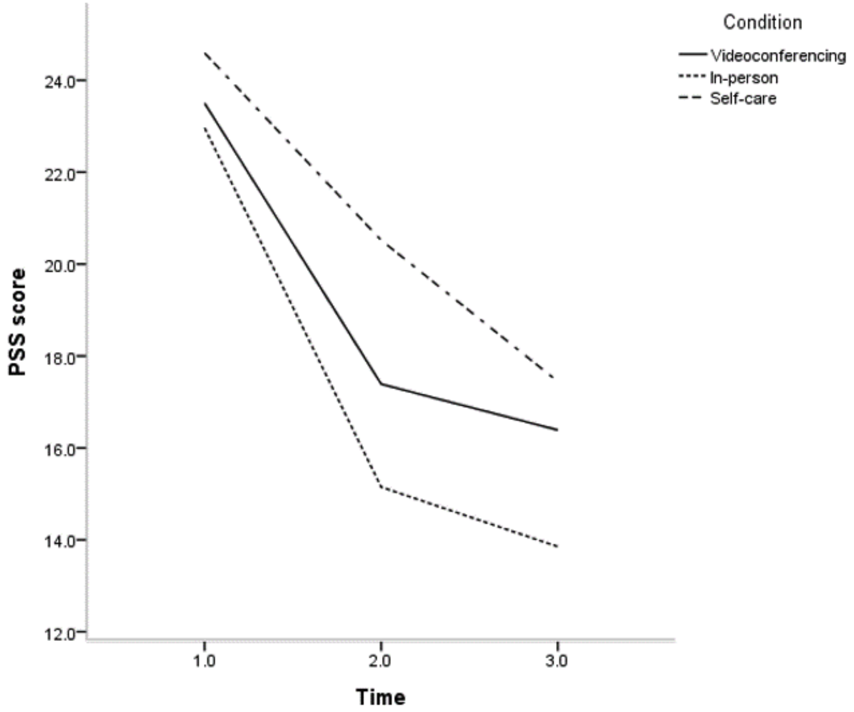


Figure 3. Change in Korean Emotional Labor Scale (KELS) scores across time according to condition.

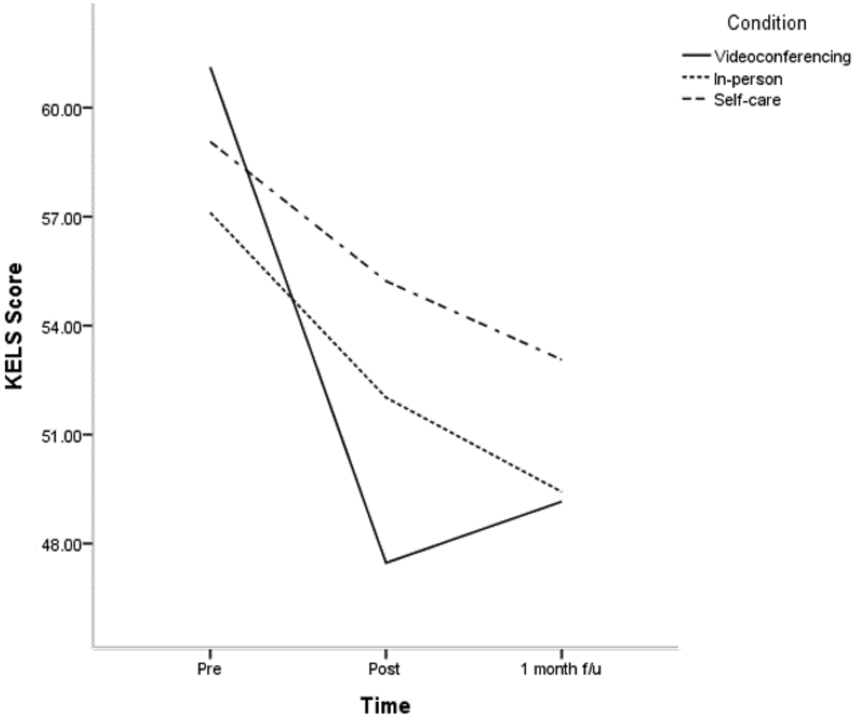


Figure 4. Change in Brief Resilience Scale (BRS) scores across time according to condition.

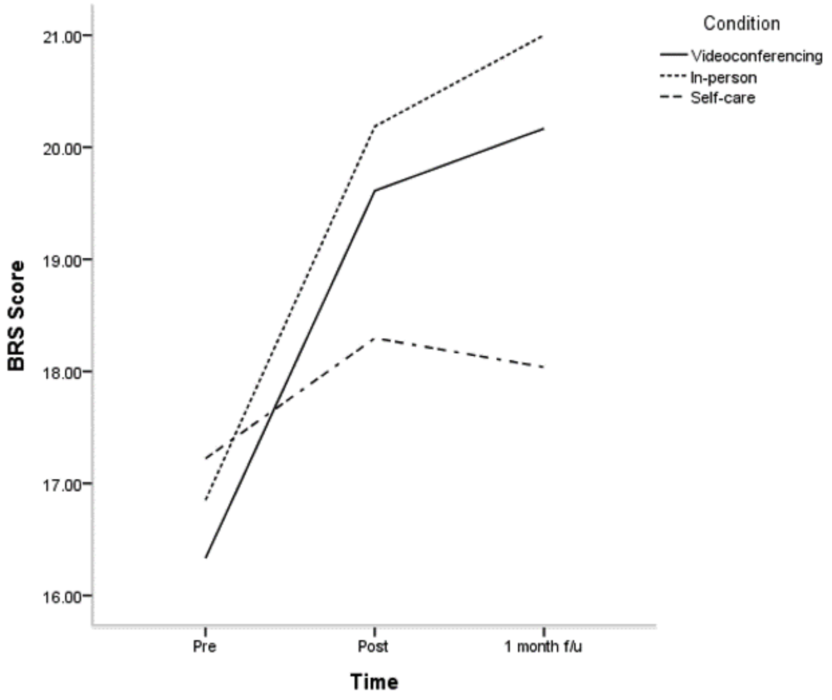
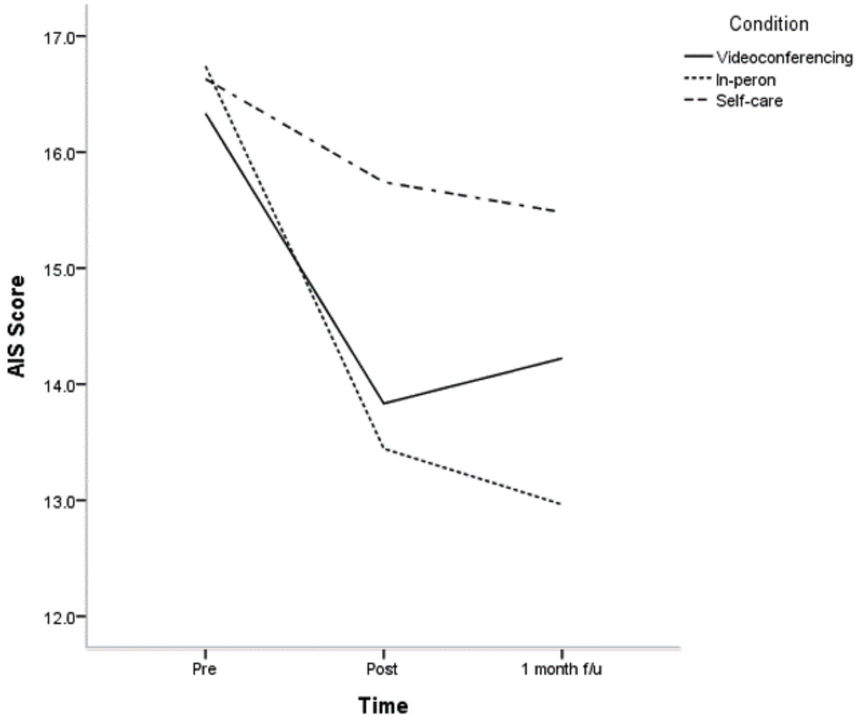


Figure 5. Change in Athens Insomnia Scale (AIS) scores across time according to condition.



Discussion

To the best of our knowledge, this is the first study investigating the effectiveness of a mobile videoconferencing-based intervention for stress reduction and resilience enhancement in employees. Videoconferencing was delivered using a smartphone app, thus combining the merits of both videoconferencing and mobile devices. Overall, there were significantly differential effects across time according to treatment condition on perceived stress, resilience, emotional labor, and sleep. Moreover, there were no significant differences between the mobile videoconferencing and in-person conditions at follow-up, suggesting that the benefits of mobile videoconferencing therapy were comparable to those of in-person therapy.

At follow-up, both the mobile videoconferencing and in-person conditions had significantly enhanced resilience compared with the self-care condition, which was the primary goal of the intervention. Resilience enhancement has been recognized as an important part of stress reduction. Resilience refers to the process that allows individuals to adapt positively despite stress or trauma [47]. Resilience-based interventions emphasize strengths within individuals and community members to persevere and recover from environmental, physical, or emotional stress [48]. Previous research regarding resilience suggests approaches that build on strengths rather than problem-focused strategies [47]. Southwick and Charney suggested 5 essential components to resilience interventions: (1) emotional regulation training to recognize and manage reactivity and impulsivity; (2) cognitive behavioral approaches to reframe thought processes and increase positive emotion; (3) physical health information on exercise, nutrition, sleep, and relaxation to increase protective behavior; (4) social support to build connections with family, peers, and mentors to increase protective factors; and (5) neurobiological components, such as mindfulness-based stress reduction, to increase the ability to manage stress [49]. Our intervention contained most of these components, leading to successful resilience enhancement.

Most studies on videoconferencing-based treatments have focused on addressing rural populations. A few studies have investigated the effectiveness of videoconferencing in homebound populations, such as disabled adults with dementia or older adults with depression [50,51]. This study suggests that telepsychology methods can also be useful to young employees living in metropolitan areas. Most adults in their 30s and 40s are smartphone savvy, which can allow them to access treatment at their convenience as long as they have a Wi-Fi or LTE (Long-Term Evolution) connection [10]. Videoconferencing via smartphones is also cost-effective as most people already own smartphones and they will be able to save travel costs [52].

To date, a variety of effective stress management programs for employees has been proposed and made available [53]. A recent meta-analysis found that while cognitive behavioral therapy interventions yielded the largest effect sizes, relaxation and meditation techniques were the most popular [54]. A SMART-3RP program that included both of these components has been found to improve resilience and minimize perceived stress in palliative care clinicians, medical interpreters, and resident physicians [44]. The intervention duration for the original SMART-3RP program was 8 weeks; however, this study showed that the intervention was effective only after 4 sessions. Moreover, this intervention was effective and also had a low drop-out rate (videoconferencing condition 14.2%, in-person condition 0% after treatment engagement) compared with previous studies. For comparison, a recent meta-analysis found that the mean completion of workplace psychological treatments was 45%, with a range of 3%-95% [55]. With respect to the 0% drop-out rate in the in-person condition, almost 40% of participants were employees of the hospitals in which the study took place; hence, high accessibility may have contributed to the low drop-out rate.

Although there was no statistical significance at follow-up, the post hoc analysis revealed that the mobile videoconferencing condition had a greater effect on emotional labor than both the in-person and self-care conditions at posttreatment. Emotional labor is a unique type of stress experienced in employees. Brotheridge and Grandey proposed that there may be two sources to job-related stress: emotional demands of the work environment and employees' ability to control their emotions. This indicates that workers with emotionally demanding jobs and a low capacity for emotional control likely experience the greatest job-related stress [29]. Of the participants who completed the study, 19 (26.4%) were nurses. Given that emotional labor is key to making patients feel safe and comfortable [56], many nurses experience emotional labor, resulting in higher job stress, poorer health, greater self-alienation, and increased frequency of depressive mood [57]. The results of this study suggest that mobile videoconference interventions may be effective in reducing emotional labor in employees, at least in the short-term.

A major concern of videoconference treatment is the quality of therapeutic alliance between patient and therapist [58]. Therapeutic alliance has been defined as collaborative effort by the therapist and patient to facilitate healing [59]. In this study, there were no significant differences in therapeutic alliance ratings between the conditions. Our findings demonstrate that stress intervention via mobile videoconferencing does not compromise therapeutic alliance, which is in line with previous telepsychology research results [60].

There are a few notable limitations to this study. The length of intervention and follow-up interval were relatively short. Thus, this study provides no information on the long-term effect of mobile videoconference interventions. Participants were mostly female and all were Korean, limiting the generalizability of the findings to the male gender and other ethnic groups. We did not exclude participants with depression, insomnia, or anxiety disorders, making the study population clinically heterogeneous; as the sample size was not sufficient for subgroup analyses according to the presence or absence of psychiatric diagnosis, we were unable to evaluate whether psychiatric diagnoses influenced the effect of mobile videoconference treatment. Moreover, we did not measure if the intervention improved any workplace variables, such as work performance, absenteeism, and turnover rate. There is also the possibility of a selection bias caused by the recruitment of highly motivated participants. Lastly, participants and therapists were not blinded to their treatment conditions, which may have caused an expectation bias.

In conclusion, this study demonstrates that videoconferencing-based stress reduction interventions can be effective in employees. Further studies with larger sample sizes and longer follow-up intervals may be helpful for determining the long-term effect of this intervention.

Acknowledgments

This research was supported from a fund by the Ministry of Trade, Industry and Energy of South Korea (No. 10069105 to J-HK).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Protocol of stress reduction and resilience enhancement intervention.
[PDF File (Adobe PDF File), 15 KB - jmir_v20i10e10760_app1.pdf]

Multimedia Appendix 2
Changes in clinical scores across time according to condition.
[PDF File (Adobe PDF File), 33 KB - jmir_v20i10e10760_app2.pdf]

Multimedia Appendix 3
CONSORT‐EHEALTH checklist (V.1.6.1).
[PDF File (Adobe PDF File), 2462 KB - jmir_v20i10e10760_app3.pdf]

References

1. Akerstedt T. Psychosocial stress and impaired sleep. Scand J Work Environ Health 2006 Dec;32(6):493-501 [FREE Full text] [Medline: 17173205]
2. Richardson S, Shaffer JA, Falzon L, Krupka D, Davidson KW, Edmondson D. Meta-analysis of perceived stress and its association with incident coronary heart disease. Am J Cardiol 2012 Dec 15;110(12):1711-1716 [FREE Full text] [doi: 10.1016/j.amjcard.2012.08.004] [Medline: 22975465]
3. Steptoe A, Kivimäki M. Stress and cardiovascular disease. Nat Rev Cardiol 2012 Apr 03;9(6):360-370. [doi: 10.1038/nrcardio.2012.45] [Medline: 22473079]
4. Hammen C. Stress and depression. Annu Rev Clin Psychol 2005;1:293-319. [doi: 10.1146/annurev.clinpsy.1.102803.143938] [Medline: 17716090]
5. Paoli P, Merllie D. Third European survey on working conditions 2000. Luxembourg: Office for official publications of the European communities; 2001.
6. Atkinson W. Stress: risk management's most serious challenge? Risk Manage 2004;51(6):20-24.
7. Rosch P. The quandary of job stress compensation. Stress Health 2001;3:1-4.
8. Bhui K, Dinos S, Stansfeld S, White P. A synthesis of the evidence for managing stress at work: a review of the reviews reporting on anxiety, depression, and absenteeism. J Environ Public Health 2012;2012:515874 [FREE Full text] [doi: 10.1155/2012/515874] [Medline: 22496705]
9. Ryan C, Bergin M, Chalder T, Wells JS. Web-based interventions for the management of stress in the workplace: Focus, form, and efficacy. J Occup Health 2017 May 25;59(3):215-236 [FREE Full text] [doi: 10.1539/joh.16-0227-RA] [Medline: 28320977]
10. 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 Jun 23;8(2):e10 [FREE Full text] [doi: 10.2196/jmir.8.2.e10] [Medline: 16867965]
11. Chakrabarti S. Usefulness of telepsychiatry: A critical evaluation of videoconferencing-based approaches. World J Psychiatry 2015 Sep 22;5(3):286-304 [FREE Full text] [doi: 10.5498/wjp.v5.i3.286] [Medline: 26425443]
12. García-Lizana F, Muñoz-Mayorga I. Telemedicine for depression: a systematic review. Perspect Psychiatr Care 2010 Apr;46(2):119-126. [doi: 10.1111/j.1744-6163.2010.00247.x] [Medline: 20377799]
13. Bouchard S, Paquin B, Payeur R, Allard M, Rivard V, Fournier T, et al. Delivering cognitive-behavior therapy for panic disorder with agoraphobia in videoconference. Telemed J E Health 2004;10(1):13-25. [doi: 10.1089/153056204773644535] [Medline: 15104911]
14. Wootton B. Remote cognitive-behavior therapy for obsessive-compulsive symptoms: A meta-analysis. Clin Psychol Rev 2016 Feb;43:103-113. [doi: 10.1016/j.cpr.2015.10.001] [Medline: 26494179]
15. Sloan D, Gallagher M, Feinstein B, Lee D, Pruneau G. Efficacy of telehealth treatments for posttraumatic stress-related symptoms: a meta-analysis. Cogn Behav Ther 2011;40(2):111-125. [doi: 10.1080/16506073.2010.550058] [Medline: 21547778]
16. Mitchell JE, Crosby RD, Wonderlich SA, Crow S, Lancaster K, Simonich H, et al. A randomized trial comparing the efficacy of cognitive-behavioral therapy for bulimia nervosa delivered via telemedicine versus face-to-face. Behav Res Ther 2008 May;46(5):581-592 [FREE Full text] [doi: 10.1016/j.brat.2008.02.004] [Medline: 18374304]

17. Drago A, Winding TN, Antypa N. Videoconferencing in psychiatry, a meta-analysis of assessment and treatment. Eur Psychiatry 2016 Dec;36:29-37. [doi: 10.1016/j.eurpsy.2016.03.007] [Medline: 27311105]
18. Korea Internet and Secury Agency. 2016 Survey on Internet Usage. Seoul, South Korea: Korea Internet and Security Agency; 2017.
19. Melchior M, Caspi A, Milne BJ, Danese A, Poulton R, Moffitt TE. Work stress precipitates depression and anxiety in young, working women and men. Psychol Med 2007 Aug;37(8):1119-1129 [FREE Full text] [doi: 10.1017/S0033291707000414] [Medline: 17407618]
20. Ebert DD, Heber E, Berking M, Riper H, Cuijpers P, Funk B, et al. Self-guided internet-based and mobile-based stress management for employees: results of a randomised controlled trial. Occup Environ Med 2016 May;73(5):315-323. [doi: 10.1136/oemed-2015-103269] [Medline: 26884049]
21. Heber E, Ebert D, Lehr D, Cuijpers P, Berking M, Nobis S, et al. The benefit of web- and computer-based interventions for sleep: a systematic review and meta-analysis. J Med Internet Res 2017;19(2):a.
22. 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 [FREE Full text] [Medline: 9881538]
23. Yoo S, Kim Y, Noh J, Oh K, Kim C, Namkoong K, et al. Validity of Korean version of the Mini Interational Neuropsychiatric Interview. Anxiety Mood 2006;2:50-55.
24. Marinaccio A, Ferrante P, Corfiati M, Di TC, Rondinone B, Bonafede M, et al. The relevance of socio-demographic and occupational variables for the assessment of work-related stress risk. BMC Public Health 2013 Dec 10;13:1157 [FREE Full text] [doi: 10.1186/1471-2458-13-1157] [Medline: 24325192]
25. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]
26. Lee E, Chung BY, Suh C, Jung J. Korean versions of the Perceived Stress Scale (PSS-14, 10 and 4): psychometric evaluation in patients with chronic disease. Scand J Caring Sci 2015 Mar;29(1):183-192. [doi: 10.1111/scs.12131] [Medline: 24660854]
27. Chang SJ, Kang DM, Kang HT, Kim SY, Kim IA, Kim JI. Application research for Korean emotional labor scale and Korean workplace violence scale. Seoul: Occupational Safety and Health Research Institute, Korean Occupational Safety & Health Agency; 2014.
28. Hochschild A. The managed heart: commercialization of human feeling. Berkeley and Los Angeles, California: University of California Press; 1983.
29. Brotheridge C, Grandey A. Emotional labor and burnout: Comparing two perspectives of “people work”. J Vocat Behav 2002;60(1):17-39.
30. Morris J, Feldman D. The Dimensions, Antecedents, and Consequences of Emotional Labor. The Academy of Management Review 1996 Oct;21(4):986-1010. [doi: 10.2307/259161]
31. Brotheridge CM, Lee RT. Testing a conservation of resources model of the dynamics of emotional labor. J Occup Health Psychol 2002 Jan;7(1):57-67. [Medline: 11827234]
32. Gilstrap C. Closing in on closeness: Teacher immediacy as a form of emotion labor. USA: Purdue University, ProQuest Dissertations Publishing; 2005.
33. Kruml S, Geddes D. Exploring the Dimensions of Emotional Labor. Management Communication Quarterly 2016 Aug 17;14(1):8-49. [doi: 10.1177/0893318900141002]
34. Blau G, Fertig J, Surges TD, Connaughton S, Soo PD, Marshall C. Further scale refinement for emotional laborxploring distinctions between types of surface versus deep acting using a difficult client referent. Career Dev Int 2010;15(2):188-216.
35. Van Dijk PA, Brown AK. Emotional labour and negative job outcomes: An evaluation of the mediating role of emotional dissonance. J Manag Organ 2006;12(2):101-115.
36. Mann S. Emotion at Work: To What Extent are We Expressing, Suppressing, or Faking It? European Journal of Work and Organizational Psychology 1999 Sep;8(3):347-369. [doi: 10.1080/135943299398221]
37. Brotheridge C, Lee R. Development and validation of the emotional labour scale. J Occup Organ Psychol 2003;76(3):365-379.
38. Best R, Downey R, Jones R. editors. 1997 Presented at: Incumbent perceptions of emotional work requirements. 12th annual conference of the Society for Industrial and Organizational Psychology; 1997; St Louis, MO.
39. Zapf D, Vogt C, Seifert C, Mertini H, Isic A. Emotion Work as a Source of Stress: The Concept and Development of an Instrument. European Journal of Work and Organizational Psychology 1999 Sep;8(3):371-400. [doi: 10.1080/135943299398230]
40. Chang S, Koh S, Kang D, Kim S, Kang M, Lee C, et al. Developing an occupational stress scale for Korean employees. Korean J Occup Environ Med 2005;17:297-317.
41. Smith B, Dalen J, Wiggins K, Tooley E, Christopher P, Bernard J. The brief resilience scale: assessing the ability to bounce back. Int J Behav Med 2008;15(3):194-200. [doi: 10.1080/10705500802222972] [Medline: 18696313]
42. Salisu I, Hashim N. A Critical Review of Scales Used in Resilience Research. IOSR JBM 2017 Apr;19(04):23-33. [doi: 10.9790/487X-1904032333]
43. Soldatos CR, Dikeos DG, Paparrigopoulos TJ. Athens Insomnia Scale: validation of an instrument based on ICD-10 criteria. J Psychosom Res 2000 Jun;48(6):555-560. [Medline: 11033374]

44. Jeong HS, Jeon Y, Ma J, Choi Y, Ban S, Lee S, et al. Validation of the Athens Insomnia Scale for screening insomnia in South Korean firefighters and rescue workers. Qual Life Res 2015 Oct;24(10):2391-2395. [doi: 10.1007/s11136-015-0986-7] [Medline: 25894062]
45. Bui E, Chad-Friedman E, Wieman S, Grasfield R, Rolfe A, Dong M, et al. Patient and Provider Perspectives on a Mind-Body Program for Grieving Older Adults. Am J Hosp Palliat Care 2018 Jun;35(6):858-865. [doi: 10.1177/1049909117743956] [Medline: 29172636]
46. https://hellomindcare.com. URL: https://hellomindcare.com/.]; [accessed 2018-09-07] [WebCite Cache ID 72FK4OkCO]
47. Luthar S, Cicchetti D, Becker B. The construct of resilience: a critical evaluation and guidelines for future work. Child Dev 2000;71(3):543-562 [FREE Full text] [Medline: 10953923]
48. Zautra A, Arewasikporn A, Davis M. Resilience: Promoting Well-Being Through Recovery, Sustainability, and Growth. Research in Human Development 2010 Aug 27;7(3):221-238. [doi: 10.1080/15427609.2010.504431]
49. Southwick SM, Charney DS. The science of resilience: implications for the prevention and treatment of depression. Science 2012 Oct 05;338(6103):79-82. [doi: 10.1126/science.1222942] [Medline: 23042887]
50. Choi NG, Hegel MT, Marti N, Marinucci ML, Sirrianni L, Bruce ML. Telehealth problem-solving therapy for depressed low-income homebound older adults. Am J Geriatr Psychiatry 2014 Mar;22(3):263-271 [FREE Full text] [doi: 10.1097/JGP.0b013e318266b356] [Medline: 23567376]
51. Tang W, Chiu H, Woo J, Hjelm M, Hui E. Telepsychiatry in psychogeriatric service: a pilot study. Int J Geriatr Psychiatry 2001 Jan;16(1):88-93. [Medline: 11180491]
52. Simpson S. Psychotherapy via videoconferencing: a review. British Journal of Guidance & Counselling 2009 Aug;37(3):271-286. [doi: 10.1080/03069880902957007]
53. van der Klink JJ, Blonk RW, Schene AH, van Dijk FJ. The benefits of interventions for work-related stress. Am J Public Health 2001 Feb;91(2):270-276. [Medline: 11211637]
54. Richardson K, Rothstein H. Effects of occupational stress management intervention programs: a meta-analysis. J Occup Health Psychol 2008 Jan;13(1):69-93. [doi: 10.1037/1076-8998.13.1.69] [Medline: 18211170]
55. Carolan S, Harris P, Cavanagh K. Improving Employee Well-Being and Effectiveness: Systematic Review and Meta-Analysis of Web-Based Psychological Interventions Delivered in the Workplace. J Med Internet Res 2017 Jul 26;19(7):e271 [FREE Full text] [doi: 10.2196/jmir.7583] [Medline: 28747293]
56. Gray B, Smith. Emotional labour and the clinical settings of nursing care: the perspectives of nurses in East London. Nurse Educ Pract 2009 Jul;9(4):253-261. [doi: 10.1016/j.nepr.2008.08.009] [Medline: 18842462]
57. Kovacs M, Kovacs E, Hegedu K. Emotion work and burnout: cross-sectional study of nurses and physicians in Hungary. Croat Med J 2010 Oct;51(5):432-442 [FREE Full text] [Medline: 20960593]
58. Simpson S, Reid C. Therapeutic alliance in videoconferencing psychotherapy: a review. Aust J Rural Health 2014 Dec;22(6):280-299. [doi: 10.1111/ajr.12149] [Medline: 25495622]
59. Bachelor A, Horvath A. The therapeutic relationship. In: Hubble MA, Duncan, B.L., Miller, S. D., editor. The heart and soul of change: what works in therapy. Washington, DC: American Psychological Association; 1999:133-178.
60. Simpson S, Bell L, Knox J, Mitchell D. Therapy via videoconferencing: a route to client empowerment? Clin. Psychol. Psychother 2005 Mar;12(2):156-165. [doi: 10.1002/cpp.436]

Abbreviations

AIS: Athens Insomnia Scale
ANOVA: analysis of variance
BRS: Brief Resilience Scale
KELS: Korean Emotional Labor Scale
KOSS: Korean Occupational Stress Scale
MINI: Mini-International Neuropsychiatric Interview
PSS: Perceived Stress Scale
SMART-3RP: Stress Management and Resilience Training: Relaxation Response Resilience Program

Edited by G Eysenbach; submitted 12.04.18; peer-reviewed by C Ryan, TR Soron; comments to author 23.05.18; revised version received 20.06.18; accepted 21.06.18; published 22.10.18.

Please cite as:
Kim JI, Yun JY, Park H, Park SY, Ahn Y, Lee H, Kim TK, Yoon S, Lee YJ, Oh S, Denninger JW, Kim BN, Kim JH
A Mobile Videoconference-Based Intervention on Stress Reduction and Resilience Enhancement in Employees: Randomized Controlled Trial
J Med Internet Res 2018;20(10):e10760
URL: http://www.jmir.org/2018/10/e10760/
doi:10.2196/10760
PMID:30348630

Original Paper

Effect of an eHealth Intervention to Reduce Sickness Absence Frequency Among Employees With Frequent Sickness Absence: Randomized Controlled Trial

Annette Notenbomer[1], MD; Corné Roelen[2], PhD; Johan Groothoff[1], PhD; Willem van Rhenen[2,3], PhD; Ute Bültmann[1], PhD

[1]Division Community and Occupational Medicine, Department of Health Sciences, University Medical Center Groningen, University of Groningen, Groningen, Netherlands

[2]Arbo Unie, Utrecht, Netherlands

[3]Business Universiteit Nyenrode, Breukelen, Netherlands

Corresponding Author:
Annette Notenbomer, MD
Division Community and Occupational Medicine, Department of Health Sciences
University Medical Center Groningen
University of Groningen
Gebouw 3217 FA10, kamer 609
Antonius van Deusinglaan 1
Groningen, 9713 AV
Netherlands
Phone: 31 6 22417514
Email: annette.notenbomer@gmail.com

Abstract

Background: Frequent sickness absence—that is, 3 or more episodes of sickness absence in 1 year—is a problem for employers and employees. Many employees who have had frequent sickness absence in a prior year also have frequent sickness absence in subsequent years: 39% in the first follow-up year and 61% within 4 years. Moreover, 19% have long-term sickness absence (≥6 weeks) in the first follow-up year and 50% within 4 years. We developed an electronic health (eHealth) intervention, consisting of fully automated feedback and advice, to use either as a stand-alone tool (eHealth intervention–only) or combined with consultation with an occupational physician (eHealth intervention–occupational physician).

Objective: This study aimed to evaluate the effect of the eHealth intervention, with or without additional occupational physician consultation, to reduce sickness absence frequency for employees with frequent sickness absence, versus care as usual (CAU).

Methods: This study was a three-armed randomized controlled trial. Employees with frequent sickness absence received invitational letters, which were distributed by their employers. The primary outcome measure was the number of register-based sickness absence episodes 12 months after completing the baseline questionnaire. Secondary outcome measures were register-based total sickness absence days and self-assessed burnout, engagement, and work ability. In a process evaluation 3 months after baseline, we examined adherence to the intervention and additional actions such as general practitioner and occupational physician visit, communication with the manager, and lifestyle change.

Results: A total of 82 participants were included in the analyses, 21 in the eHealth intervention–only group, 31 in the eHealth intervention–occupational physician group, and 30 in the CAU group. We found no significant difference in sickness absence frequency between the groups at 1-year follow-up. Sickness absence frequency decreased in the eHealth intervention–only group from 3 (interquartile range, IQR 3-4) to 1 episode (IQR 0.3-2.8), in the eHealth intervention–occupational physician group from 4 (IQR 3-5) to 3 episodes (IQR 1-4), and in the CAU group from 3 (IQR 3-4) to 2 episodes (IQR 1-3). For secondary outcomes, we found no significant differences between the intervention groups and the control group. The process evaluation showed that only 3 participants from the eHealth intervention–occupational physician group visited the occupational physician on invitation.

Conclusions: Among employees with frequent sickness absence, we found no effect from the eHealth intervention as a stand-alone tool in reducing sickness absence frequency, nor on total sickness absence days, burnout, engagement, or work ability. This might be due to low adherence to the intervention because of insufficient urgency to act. We cannot draw any conclusion on the effect of the eHealth intervention tool combined with an occupational physician consultation (eHealth intervention–occupational physician), due to very low adherence to the occupational physician consultation. An occupational physician consultation could increase a sense of urgency and lead to more focus and appropriate support. As this was the first effectiveness study among employees with frequent sickness absence, strategies to improve recruitment and adherence in occupational eHealth are included.

Trial Registration: Netherlands Trial Register NTR4316; http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=4316 (Archived by WebCite at http://www.webcitation.org/713DHhOFU).

(J Med Internet Res 2018;20(10):e10821) doi:10.2196/10821

KEYWORDS

occupational health; sick leave; randomized controlled trial; adult; occupational health physicians; eHealth

Introduction

Background

Frequent sickness absence (SA)—that is, 3 or more episodes of SA in a year,—is common in the Dutch working population and poses a problem for both employers and employees [1]. The prevalence of frequent SA was 5.8% in 2015 and 6.1% in 2016 among more than 600,000 employees working in small- to medium-sized companies under contract with a Dutch national occupational health service (OHS). In the Netherlands, the costs related to frequent SA can be estimated to be at least 100 million euros per year (US $123 million) [2]. Frequent SA is often a precursor of future frequent SA; Koopmans et al found that 39% of employees with frequent SA had frequent SA again in the following year and 61% within 4 years [3]. In addition, there is a risk of high future costs due to long-term SA, as 19% of frequent absentees have long-term SA (≥6 weeks) in the first follow-up year and 50% in the 4 years following a year with frequent SA [3]. Employers have to redirect work tasks to other employees and are faced with work planning problems. Employees with frequent SA can suffer from increasing tensions with colleagues [4] who are burdened with an increased workload and overwork. Furthermore, frequent absentees are at increased risk of being dismissed [5]. In a literature review, Beemsterboer et al [6] found that poorer health, poorer working conditions, greater physical and mental workload, female gender, greater alcohol consumption, and smoking were related to a higher SA frequency. In contrast, better job resources, better working relations, greater motivation (work pleasure), older age, being married, having a satisfactory private life, and higher education were associated with a lower SA frequency. Other studies found that poor health, chronic diseases, and reduced work ability were related to frequent SA [7-10]. On the basis of a focus group study among frequent absentees, we reported earlier that high job demands and low job resources, particularly low support from management, were related to frequent SA [11]. Frequent absentees also mentioned home demands, poor health, chronic illness, and unhealthy lifestyles to be related to frequent SA [11]. The results of these studies suggest that frequent SA is a multifaceted phenomenon.

To reduce SA frequency among employees with frequent SA, an intervention is needed to address this wide range of issues. To reduce frequent SA and feel better, employees with frequent SA prefer to seek adequate (medical) help themselves [11]. They want help only when self-management fails [11]. Among published interventions to reduce SA, several studies showed successful reduction of SA in employees at high risk of future SA through face-to-face structured consultations with occupational health professionals [12-15]. However, consultations can be time-consuming and costly. In contrast, electronic health (eHealth) interventions are low in cost and appeal to self-management. The number of studies on eHealth interventions (EHIs) has grown rapidly in the last few years. Studies among patient or general populations have shown that EHIs enhance health and well-being, improve lifestyle, and increase self-management for a broad range of diseases or risk factors [16-21]. In occupational health, however, studies on EHIs are relatively scarce. eHealth studies among employees focus mainly on lifestyle change, stress reduction, or mental health improvement. Studies on effectiveness of EHIs on absenteeism measures are few. In a randomized controlled trial (RCT) among employees at high risk of SA due to depression, Beiwinkel et al tested a 12-week eHealth program and compared the results with those of a waiting-list control group that received written psychoeducation. Although both groups showed a high reduction in SA frequency (intervention group: 67% and control group: 83%), there was no statistical difference in SA frequency between the groups [22]. Ebert et al [23] found a reduction of depressive symptoms among a group of teachers through an internet-based problem-solving training but no difference between the intervention and control groups on the secondary outcome measure self-rated absenteeism. Other studies showed that a self-guided internet-based stress management intervention resulted in a significant reduction of perceived stress in a group with increased stress but no reduction in absenteeism [24-26]. Volker et al [27] reported an earlier return to work in the intervention group among sick-listed employees with common mental disorders (hazard ratio 1.39, 95% CI 1.03-1.87). During a 1-year follow-up period, the intervention group received a blended EHI, compared with the care as usual (CAU) group. The blended EHI consisted of a mix of personalized e-modules in combination with guidance by a specially trained occupational physician (OP). All these EHIs have the use of questionnaires in common for a targeted intervention and elements of self-management.

To our knowledge, no EHIs have as yet been designed to reduce SA frequency among frequent absentees. We developed an EHI tool to advise employees with frequent SA as to how they could improve health and self-management. The theoretical framework for this EHI tool was based on the Job Demands-Resources (JD-R) model [28]. The JD-R model relates psychosocial work characteristics to outcome measures such as burnout, engagement, productivity, and SA [29-32] and provides keys for guidance and support [11,32]. We used tools from the mental health guidelines of the Netherlands Society for Occupational Physicians (NVAB) [33]. Furthermore, the EHI tool was based on the determinants of frequent SA according to focus group participants and their suggestions on how to reduce frequent SA, such as communication with management [11]. The intervention addressed these elements item by item.

In the Netherlands, OPs often play a role in work-related interventions to reduce SA as they are experts in health, work, SA, and prevention. They can also address factors from the JD-R model, advising (temporary) accommodations to reduce job demands or increase job resources. The role of OPs in effective reduction of SA [12,14,15] led us to include a blended-care study arm, combining the EHI tool with a consultation with the OP.

Objective

The main objective of this study was to evaluate the effectiveness of the EHI tool without (EHI-only) and with OP consultation (EHI-OP), compared with CAU, on SA frequency among employees with frequent SA. Secondary outcomes were the total number of SA days, burnout, work engagement, and work ability at 1-year follow-up. We conducted a process evaluation to evaluate adherence to the intervention.

Methods

Design and Participants

This study was designed as a three-armed RCT and registered in the Dutch trial register (NTR 4136). The Medical Ethics Committee (METc) of the University Medical Center Groningen approved the study (METc 2013/131). Employees participated voluntarily and signed an informed consent form (see Multimedia Appendix 1). The report is based on the CONSORT eHealth checklist [34] (see Multimedia Appendix 2).

Inclusion criteria were employees with frequent SA, that is, 3 or more episodes of SA in the year before recruitment, irrespective of the causes or duration of sick leave. Exclusion criterion was inability to complete a questionnaire in Dutch.

Study participants were recruited from 21 Dutch organizations staffing more than 100 employees between December 2013 and December 2014. Of the participating organizations, 7 were in industry, 5 in commercial services, and 9 in public services. The first author (AN) had prepared a list of all frequent absentees in the participating organizations (source population), based on the occupational health service register. All employees with frequent SA received from us invitational letters combined with informed consent forms, which were distributed by their employers. The letters contained logos from both the University of Groningen and the OHS. Upon signing the informed consent form, we sent a personal URL code, which provided access to the Web-based questionnaire at baseline (T0). The questionnaire included validated scales measuring secondary outcome measures such as burnout, engagement, and work ability as well as potential determinants of frequent SA as found in the focus group study: job demands, job resources, health, chronic diseases, lifestyle aspects, and a few open questions on health and frequent SA.

Upon completion of the baseline questionnaire, the intervention groups received fully automated personal advice. The control group was thanked for participation upon completion of the questionnaire. All participants were invited to fill out a process evaluation 3 months after T0 and a short questionnaire 1 year after baseline (T1), with questions on the outcome measures.

Intervention

Intervention Group 1: eHealth Intervention Only (EHI-only)

The main scope of the intervention was self-management with help from relevant others, for example, the employer, OP, and general practitioner (GP). The intervention consisted of immediate fully automated personalized Web-based feedback, item by item. The addressed items were job demands (work pace, emotional demands, and work-home interference), job resources (feedback, learning opportunities, supervisor support, coworker support, and autonomy), burnout, engagement, work ability, general health, chronic diseases, psychological health, lifestyle, and body mass index. For an overview of the intervention elements per item, see Multimedia Appendix 3. The feedback per item consisted of the score, interpretation of the score, general advice on possibilities to tackle this issue (in case of a poor score), reference to relevant internet sites for more information, further diagnostic tests or treatment, and referral to people who could help (depending on the issue: manager, colleagues, human resource management [HRM], OP, social worker, or GP). The advice often contained a link to documents with more detailed advice. The advice was based on NVAB guidelines, occupational health care practice, and suggestions from focus group participants with frequent SA from a prior study [11]. Cut-off points were either the existing cut-off points of the scales or the seventy-fifth percentile of a large reference group who participated in OHS health surveillance checks. Participants who scored well on a particular scale received feedback that they had scored well and no specific actions were needed.

Intervention Group 2: eHealth Intervention + Occupational Physician Consultation (EHI-OP)

Participants in intervention group 2 received the same advice and documents as the EHI-only group but were invited by email to a preventive advisory consultation with the OP. The email contained the name of the OP and the telephone number of the OP’s secretary to make an appointment.

The OPs from the 21 participating organizations received written information on the study (see Multimedia Appendix 4) and a personal explanation by the first author (AN) about the goal of the study and the possibility of consultations with participants. Moreover, AN explained that what was expected in this preventive consultation was the same as in preventive consultations initiated by the employee in nonresearch situations: mainly participants’ questions on health and SA in relation to work and how to influence the employee’s health or (work) situation. This could lead to making a joint plan-of-action, but it was not obligatory. Standard time for this preventive consultation was 30 min.

Control Group: Care as Usual

The control group received neither personalized advice nor support from the OP or researchers upon completion of the Web-based questionnaire. CAU consisted of consultation with the OP at the request of the employer or control group participant. In case of long-term SA, participants were invited for a consultation with the OP to certify SA within 6 weeks of reporting sick [35].

Pilot Test of eHealth Intervention

We pilot tested the EHI tool in 12 frequent absentees from 3 nonparticipating organizations. We used their feedback on technical issues and understandability to improve the EHI tool. After finalization of the tool, we made no changes in the contents.

Primary Outcome Variable

The primary outcome measure was the number of register-based SA episodes 12 months after completing the Web-based baseline questionnaire. At 1-year follow-up, the incident number of SA episodes was retrieved at the individual level from the occupational health service register, in which SA was recorded from the first day of sick leave to the day of return to work.

Secondary Outcome Variables

The number of days of all SA episodes was cumulated to a total number of SA days at 1-year follow-up. Burnout, work engagement, and work ability were measured at 1-year follow-up.

Burnout was measured with the 9-item Utrecht Burnout Scale (UBOS) measuring emotional exhaustion and cynicism on a 7-point frequency scale ranging between never (=0) and always (=6) [36]. A total UBOS-9 score (Cronbach alpha=.92) was calculated by summing the item scores, with higher scores representing higher levels of burnout. The scale included the dimensions emotional exhaustion (5 items) and cynicism (4 items).

Work engagement was measured with the 9-item Utrecht Work Engagement Scale (UWES) [37]. The UWES scale's vigor (3 items), dedication (3 items), and absorption (3 items) were scored on a 7-point frequency scale ranging from “never” (=0) to “always” (=6). A total UWES-9 score was calculated by summing the item scores (Cronbach alpha=. 95), with higher scores representing more work engagement.

Work ability was investigated with the first item of the Work Ability Index (WAI), asking for current work ability compared with lifetime best on a 0 to 10 scale. This single-question work ability score has been reported to be a reasonable alternative to the complete WAI for measuring work ability in working populations [38,39].

Work and Population Characteristics

We assessed the following items to provide targeted intervention advice. Job demands (work pace, emotional demands, and work-home interference) and job resources (feedback, learning opportunities, supervisor support, coworker support, and autonomy) were measured with the short scales of the Questionnaire on the Experience and Evaluation of Work [40].

General health was measured with the Short-Form Health Survey, the single-item question: “In general, would you say your health is excellent (=5), very good (=4), good (=3), fair (=2), or poor (=1)?” This item has been associated with physicians’ assessments of health, morbidity measures, and utilization of health services [41] and is comparable with longer instruments [42]. The presence of chronic diseases was assessed with the item “Do you have a chronic disease that already lasts >3 months?” (yes/no). The Alcohol Use Disorder Identification Test [43] was used to measure alcohol consumption. Physical exercise was measured with Dutch Norm Healthy Moving [44] and Fitnorm [45]. Smoking was assessed by a single-item question on smoking (yes/no) and relaxation with 2 single-item questions: “Do you have at least half an hour relaxation every day?” and “Do you regularly take a break at work?” Response options were “hardly every/never,” “regularly,” and “(almost) always.” Additionally included were the questions “Do you find it a problem that you are frequently on sick-leave?” (yes, no, do not know), “Does your supervisor find it a problem that you are frequently on sick-leave?” (yes, no, do not know), and “Do your colleagues find it a problem that you are frequently on sick-leave?” (yes, no, do not know, some do/some do not). An open question at the end of the questionnaire requested other relevant information (“What else plays a role in your health and SA that has not yet been covered?”).

Process Evaluation

In a process evaluation 3 months after baseline, adherence to the intervention was measured in the EHI groups, assessing reading the advice provided by the eHealth tool and undertaking actions. Actions assessed were a consultation with the OP, GP, specialists, paramedics, or psychologist and additional actions such as tackling sources of stress, tackling work-related problems, and having a conversation at work about work-related problems or solutions, lifestyle changes, and other actions (open question). Participants from the control group also received a process evaluation, with questions on OP consultation and additional actions, for example, visit to GP, physiotherapist, psychologist, or other paramedics; lifestyle changes; or consultations with management, HRM, or occupational health providers to seek solutions for work-related problems.

Sample Size

In a pilot study, we found that frequent absentees had on average 3.79 (SD 1.27) SA episodes in 2013 to 2014 in the total employee population of a large Dutch OHS. RCT intervention studies that include SA frequency as an outcome measure are scarce. No scientifically based intervention effect was available as this was the first intervention study among employees with frequent SA on SA frequency. The RCT studies from Duijts et al [13] and Kant et al [15] are the closest scientific approaches to this intervention study, although targeted at a different population. Applying their results to our study on frequent absentees, we aimed in the original trial protocol for a reduction of 0.5 episodes (Cohen d=0.39) with our focused intervention. On the basis of an alpha of .05 (two-tailed) and a power of 80%, a sample size of 103 was needed [46]. After further consideration, we included in the submission to the METc, before the start of the study, a second sample size calculation to detect a difference of 1 SA episode per year (Cohen *d*=0.79). This was based on our practice-based knowledge of relevant intervention effects in an occupational health setting. This calculation showed that we needed a minimum of 27 per group [46].

Randomization

The source population (N=825) was prerandomized into 3 arms: intervention group 1 (EHI-only; n=270), intervention group 2 (EHI-OP; n=279), and control group (n=276) by random integers [47]. We randomized the entire source population as the software provider was only able to generate and provide URLs groupwise.

Blinding

Participants were allocated to the intervention groups and control group before the study started. They were blinded for the group to which they were allocated until completion of the Web-based questionnaire, whereupon they did (intervention groups) or did not (control group) receive a personalized advice. The first author (AN) knew to which group each individual belonged. SA data were retrieved and analyzed by another author (CR) who did not know to which group each individual belonged.

Statistical Analyses

All statistical analyses were conducted in IBM SPSS Statistics for Windows, version 24 (released 2016; IBM Corp. Armonk, NY). Baseline data on primary and secondary outcomes were register-based or provided by participants. Missing information from follow-up assessments was imputed using baseline observation carried forward.

Analysis of 3 Study Arms

First, we investigated differences in outcomes at T1 between the EHI-only group, the EHI-OP group, and the CAU group according to the intention-to-treat principle. Due to the non-normal distribution of incident SA episodes and days, we investigated differences by using the nonparametric Kruskal-Wallis test.

Analysis of Combined Intervention Groups Versus Control Group

The intervention groups (EHI-only and EHI-OP) were merged, as all participants from these study arms had access to the same EHI and only 3 (13%) participants from the EHI-OP group additionally consulted the OP upon invitation. We investigated the differences between the outcomes of the combined intervention groups and the control group by using the nonparametric Mann-Whitney U test.

Results

Descriptive Statistics

All 825 eligible employees with frequent SA from 21 participating organizations (the source population) were randomized. A total of 2 reorganizing organizations withdrew from the study after receiving negative reactions from their employees (n=122 eligible employees). In 6 other organizations, none of the eligible employees (n=163) decided to participate. From the other 13 organizations, 525 employees did not send an informed consent form. Some employees (n=15) filled out the consent form but failed to complete the baseline questionnaire. One employee was misregistered as having frequent SA. After exclusion of these 743 eligible employees, the study included 82 participants at baseline. SA-registered data of 3 participants were lost to follow-up due to temporary contracts that ended during the follow-up period. In total, 17 participants did not fill out the last questionnaire. Finally, 21 participants were included in the EHI-only group, 31 in the EHI-OP group, and 30 in the CAU group in the analysis. Figure 1 provides an overview of the recruitment flow.

Figure 1. Flowchart of participants. CAU: care as usual; EHI-only: study arm with only eHealth intervention; EHI-OP: study arm with eHealth intervention and invitation for OP consultation.

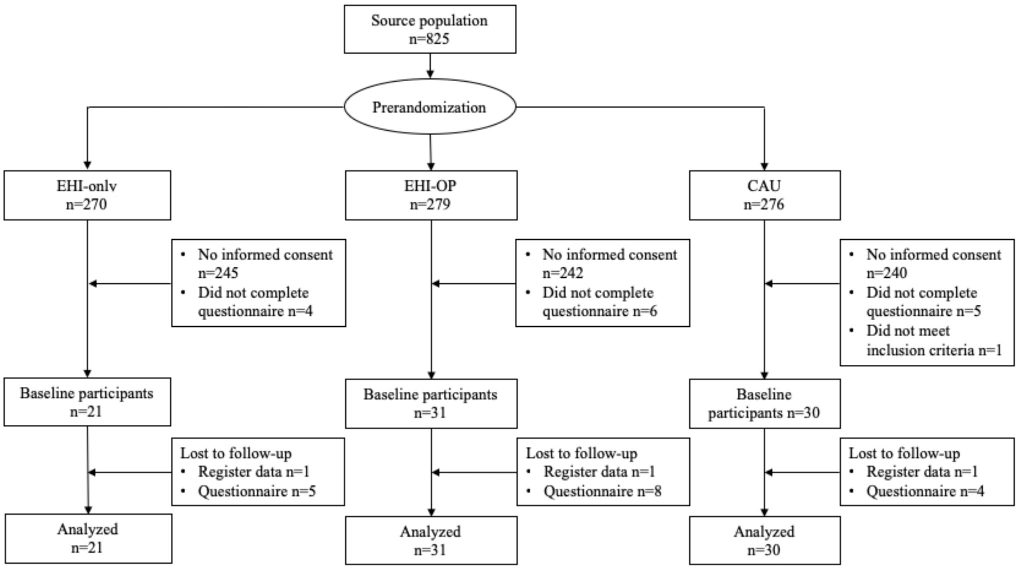


Table 1. Characteristics of participants.

Participant characteristics	EHI-only[a] (N=21)	EHI-OP[b] (N=31)	CAU[c] (N=30)
Sociodemographic characteristics			
Age (years), mean (SD)	44.9 (10.1)	45.9 (11.4)	46.9 (10.9)
Gender (male), n (%)	7 (33)	10 (32)	10 (33)
Marital status (married or living together), n (%)	18 (86)	23 (74)	21 (70)
Educational level, n (%)			
Low	4 (19)	3 (10)	2 (7)
Intermediate	9 (43)	10 (32)	11 37)
High	8 (38)	18 (58)	17 (57)
Work-related characteristics			
Irregular work (eg, night shift), n (%)	5 (24)	6 (19)	5 (17)
Years with current employer, mean (SD)	13. (8.9)	14.4 (10.9)	15.1 (11.6)
Hours per week, mean (SD)	31.1 (7.4)	34.9 (8.7)	33.1 (11.1)
Workload, mean (SD)	2.5 (0.6)	2.7 (0.7)	2.5 (0.9)
Emotional demands, mean (SD)	2.2 (1.0)	1.9 (0.8)	1.9 (0.6)
Conflict at work (current), n (%)	3 (14)	1 (3)	3 (10)
Work-home interference, mean (SD)	1.5 (0.4)	1.8 (0.8)	1.7 (0.8)
Performance feedback, mean (SD)	3.1 (1.1)	3.1 (1.0)	3.1 (1.0)
Opportunities to learn and develop, mean (SD)	2.7 (1.0)	2.6 (1.0)	2.6 (1.1)
Support manager, mean (SD)	2.9 (1.1)	2.8 (1.2)	3.1 (1.1)
Support colleagues, mean (SD)	3.5 (0.8)	3.7 (1.0)	3.4 (1.2)
Autonomy, mean (SD)	3.3 (1.0)	3.7 (0.9)	3.1 (1.0)
Health related, n (%)			
Health perception (short form health survey 1-item [SF-1])			
Bad	0 (0)	3 (10)	1 (3)
Fair	4 (19)	7 (23)	14 (47)
Good	14 (67)	14 (45)	11 (37)
Very good	2 (10)	5 (16)	3 (10)
Excellent	1 (5)	2 (7)	1 (3)
Chronic disease	7 (35)	12 (40)	8 (28)

[a] EHI-only: study arm with only eHealth intervention.

[b] EHI-OP: study arm with eHealth intervention and invitation for occupational physician consultation.

[c] CAU: care as usual.

Table 1 shows the characteristics of the 82 analyzed participants. The mean age was 46 years. The percentage of men was 33% (27/82)and 76% (62/82) were married or living together. Few participants had low education: 19% (4/21) of the EHI-only group, 10% (3/31) of the EHI-OP group, and 7% (2/30) of the CAU group. Many participants found frequent SA a problem for themselves (58/82, 71%), but a very low percentage reported that they thought that this was the case for their managers (15/82, 18%) or colleagues (11/82, 13%). The information on self-reported factors playing a role in health and frequent SA could be categorized into 5 main categories: type of disease (eg, chronic disease and migraine, 28/82, 34%), high job demands (12/82, 15%), low job resources (4/82, 5%), home demands (7/82, 9%), and imbalance between demands and capacity (4/82, 5%). An additional 26 participants (32%) did not answer this open question.

Analysis of 3 Study Arms

Primary Outcome

There was no significant difference in reduction of SA frequency between the 3 study arms (Kruskal-Wallis: *P*=.66). All 3 groups, EHI-only, EHI-OP, and CAU, showed a significant reduction in SA frequency over time (*P* values of respective Wilcoxon signed rank tests: EHI-only: *P*=.006, EHI-OP: *P*<.001, and control group: *P*<.001). Where all participants had frequent SA at baseline, at 1-year follow-up, 5 participants in the EHI-only group (5/21, 25%) had frequent SA, 16 participants in the EHI-OP group (16/31, 52%), and 12 in the CAU group (12/30, 40%, data not shown). Table 2 shows that there was no significant difference in SA frequency between the EHI groups and the CAU group at 1-year follow-up.

Secondary Outcomes

All 3 groups showed a reduction in total SA days over time. No significant difference was found between the EHI and CAU groups in the total number of SA days at 1-year follow-up (Table 2). In the EHI-only group, 3 (15%) had long-term SA (ie, ≥42 consecutive days) during 1-year follow-up, 7 in the EHI-OP group (23%), and 8 in the CAU group (28%, data not shown). We found no significant differences between the EHI groups and CAU group in burnout and engagement. Work ability was significantly lower in the EHI-OP group at 1-year follow-up compared with the other groups.

Analysis of Combined Intervention Groups Versus Control Group

Primary Outcome

The combined intervention groups (EHI groups) and the CAU group showed a reduction in SA frequency over time (Table 3). At 1-year follow-up, 21 participants (42%) in the combined EHI groups again had frequent SA as well as 12 in the CAU group (41%, data not shown). Table 3 shows that there was no significant difference in SA frequency between the EHI groups and the CAU group at 1-year follow-up.

Secondary Outcomes

The combined intervention groups and the control group showed a reduction in total SA days over time. There was no significant difference between the EHI groups and the CAU group in the total number of SA days at 1-year follow-up (Table 3). In the EHI groups, 10 participants (20%) had long-term SA (ie, ≥42 consecutive days) during 1-year follow-up, compared with 8 in the CAU group (40%, data not shown). No significant differences were found between the EHI groups and CAU group in burnout, engagement, or work ability.

Table 2. Results of 3 study arms.

Outcome at 1-year follow-up	Median T0 (IQR[a] 25-75%)	Median T1 (IQR 25-75%)	*P* value (Kruskal-Wallis)[b]
Frequency of sickness absence episodes			.43
EHI-only[c]	3 (3-4)	1 (0.5-3.5)	
EHI-OP[d]	4 (3-5)	3 (1-4)	
CAU[e]	3 (3-4)	2 (1-3)	
Total number of sickness absence days			.15
EHI-only	22 (14.5-37.5)	5 (1-25)	
EHI-OP	17 (8-34)	11 (4-36)	
CAU	20.5 (11.5-38.8)	12.5 (7.0-73.5)	
Burnout (range 0-6)			.30
EHI-only	0.9 (0.3-1.5)	1.2 (0.2-1.6)	
EHI-OP	1.4 (0.6-2.0)	1.3 (0.5-2.3)	
CAU	1.3 (0.5-2.2)	1.4 (0.8-2.2)	
Engagement (range 0-6)			.27
EHI-only	3.6 (2.4-4.5)	3.9 (2.8-4.8)	
EHI-OP	3.7 (2.6-4.7)	3.4 (2.6-4.1)	
CAU	2.8 (2.1-4.3)	3.3 (2.4-3.9)	
Work ability (range 0-10)			.01
EHI-only	8 (7-9)	8 (8-9)	
EHI-OP	8 (8-9)	7 (7-8)	
CAU	8 (7-9)	7.5 (7-8)	

[a] IQR: interquartile range.

[b] *P* values for differences between the 3 study arms at T1.

[c] EHI-only: study arm with only eHealth intervention.

[d] EHI-OP: study arm with eHealth intervention and invitation for occupational physician consultation.

[e] CAU: care as usual.

Table 3. Results of analysis with combined intervention groups versus control group.

Outcome at 1-year follow-up	Median T0 (IQR[a] 25-75%)	Median T1 (IQR 25-75%)	*P* value (Mann-Whitney U)[b]
Frequency of sickness absence episodes			.91
EHI groups[c]	3.5 (3-4.8)	2.4 (1-4)	
CAU[d]	3 (3-4)	2 (1-3)	
Total number of sickness absence days			.19
EHI groups	19 (10.3-37)	8.7 (2.3-31.5)	
CAU	20.5 (11.5-38.8)	12.5 (7-73.5)	
Burnout (range 0-6)			.29
EHI groups	1.2 (0.5-1.9)	1.3 (0.4-2.1)	
CAU	1.3 (0.5-2.2)	1.4 (0.8-2.2)	
Engagement (range 0-6)			.16
EHI groups	3.6 (2.6-4.6)	3.6 (2.6 -4.4)	
CAU	2.8 (2.1-4.3)	3.3 (2.4-3.9)	
Work ability (range 0-10)			.23
EHI groups	8 (7-9)	7.4 (7-9)	
CAU	8 (7-9)	7.5 (7-8)	

[a] IQR: interquartile range.

[b] *P* values for the combined eHealth intervention groups versus the care as usual group at T1.

[c] EHI groups: combined eHealth intervention groups.

[d] CAU: care as usual.

Process Evaluation: Adherence to the Intervention and Additional Actions

In a process evaluation 3 months after baseline, all participants received a questionnaire focusing on adherence to the intervention and additional actions they had taken. In total, 55 (70%) participants responded to the process evaluation.

A total of 3 participants out of 30 (10%) in the EHI-OP group reported that they had consulted the OP upon study invitation. Moreover, 2 participants from this study arm had seen the OP at a later moment (at the initiative of their employer) because of longer SA. A total of 2 participants from the EHI-only group (10%) and 3 from the control group (10%) reported having consulted the OP. All participants who had visited the OP were satisfied with the consultation.

In the EHI-only group, 54% (7/13) process evaluation responders reported that they had received the digital advice and 72% (13/18) in the EHI-OP group. Of the 7 receivers in the EHI-only group, 4 (57%) had read the additional documents provided. This was the case for 46% (6/13) receivers in the EHI-OP group. No participant made a plan-of-action as provided in the digital tool. For a schematic overview of adherence to the intervention and additional actions, see Table 4.

Participants from all groups—intervention and control—took additional actions. A total of 16 participants (29%) took action at work (4/13 in the EHI-only group, 31%; 5/18 in the EHI-OP group, 28%; and 7/24 in the control group, 29%). Overall, we observed no marked differences between actions in the EHI groups and the CAU group. A total of 4 participants from the EHI-only group (31%) and 9 from the EHI-OP group (50%) had taken no (new) actions since study participation. Many participants reported why they had not undertaken any further actions, including OP consultations, for example, having already taken a lot of actions before the study, still undertaking actions they had started before the study, not acknowledging the added value of the intervention when knowing the problem is not work-related, being too busy, low urgency, and optimism about their future health and SA.

Table 4. Overview of actions per group according to process evaluation.

Actions undertaken	EHI-only[a] (n=13)	EHI-OP[b] (n=18)	CAU[c] (n=24)
Participation rate in process evaluation, 3 months after intervention, %	65	60	83
Digital scores and advice received, n	7	13	N/A[d]
Digital advice, n			N/A
Fully read	7	9	
Partially read	N/A	3	
Not read	N/A	1	
Digital documents read, n	4	6	N/A
Visits to occupational physician according to process evaluation, n	2	5	3
Visit to general practitioner or medical specialist, n	1	4	3
Visit to psychologist, n	1	1	1
Tackle sources of stress, n	2	0	2
Tackle or discuss problems and solutions at work, n	4	5	7
Change lifestyle, n	3	5	7

[a] EHI-only: study arm with only eHealth intervention.

[b] EHI-OP: study arm with eHealth intervention and invitation for OP consultation.

[c] CAU: care as usual.

[d] N/A: not applicable.

Discussion

Principal Findings

The primary aim of this study was to evaluate the effectiveness of the EHI tool without OP consultation (EHI-only) or with OP consultation (EHI-OP) on SA frequency among employees with frequent SA compared with CAU. The secondary aim was to evaluate the effectiveness of the interventions (EHI groups) on the total number of SA days, burnout, engagement, and work ability. There was no significant difference in SA frequency during follow-up between the EHI groups and the CAU group. SA frequency was lower at T1 compared with T0 for all groups. We also found no significant difference on total SA, burnout, and engagement at T1 between the EHI groups and CAU group. Work ability was lower in the EHI-OP study arm than in the CAU group at 1-year follow-up in the “analysis of 3 study arms, but not in the analysis comparing the combined intervention groups versus the control group.” The combination of the EHI tool with an OP consultation was not tested, as only 3 people consulted the OP upon invitation. The results from the EHI-OP group represent more the effect of the EHI intervention as a stand-alone tool than the intended blended care.

Our findings on SA frequency are in line with previous results of RCTs on the effect of non-EHIs. Duijts et al [13] reported that a coaching intervention (7 to 9 sessions of 1 hour) did not result in a significant difference of SA episodes in the intervention group compared with the control group. However, the number of SA days at follow-up was significantly lower. Likewise, Kant et al [15] found no significant differences in SA episodes between intervention and control groups in an intervention with structured early consultations with the OP. They found a significantly lower SA duration in the follow-up period in a per-protocol analysis, but not in the intention-to-treat analyses. Although the intervention groups in our study also showed lower total SA days at 1-year follow-up compared with the control group, we found no significant difference between the groups.

That no significant differences were found between groups might be explained by too few participants in the intervention and control groups. Another explanation for finding no effect is that we may have included participants with more serious conditions than in the general population of frequent absentees, that is, conditions that are more difficult or impossible to solve or treat. One indication of having a selection of frequent absentees with more severe conditions is that some participants joined the study after having started a period of long-term SA. Another indication of selection of frequent absentees with more severe conditions is the incidence of long-term SA. In our study, 28% (8/30) in the control group had long-term SA (≥42 days), as opposed to 19% in an earlier longitudinal study on the risk of long-term SA in 4126 frequent absentees [2]. Possibly, the use of the logo of an OHS may have contributed to a selection bias toward participants with more severe conditions as OHS in the Netherlands mostly focus on employees with long-term SA.

The EHI-OP intervention might have been more effective if more frequent absentees had visited the OP for preventive advisory consultations. The resulting increased awareness of the high-risk of future long-term SA could have motivated absentees to undertake more actions to improve their situation. The OP could also have helped them to prioritize these actions by working with them to develop a structured plan of action. Moreover, the OP could have referred them to other professionals, such as a company social worker or psychologist.

According to a recent meta-analysis by Heber et al [48], guided computer-based interventions, combining computer based-intervention with human written guidance, may be more effective than stand-alone EHIs. In meta-analyses, Hutchesson et al found effectiveness of EHIs to be enhanced by the addition of face-to-face group or individual sessions or extra technologies such as self-monitoring tools, email counseling, or online group discussions [49]. Reasons for frequent absentees not to accept the invitation to consult with an OP included mistrust, insufficient urgency to take (additional) action, not acknowledging the added value of the OP when knowing the (private) problem, being too busy, and optimism about their future health and SA.

Another explanation for our results may be low adherence to the EHI tool itself; in the process evaluation, 15 participants (15/55, 27%) reported not having received or (fully) read the digital advice. Only a few participants read the additional documents. Many (23/55, 42%) did not take additional actions such as seeking advice from general practitioners, paramedics, or discussions with the employer about work problems. Although 71% (58/82) of the participants considered their frequent SA to be a problem, only 18% (15/82) of participants considered frequent SA to be a problem for the manager and only 13% (11/82) for (some of) the colleagues. Overall, it seems that frequent absentees did not feel a sense of urgency, either intrinsically or in response to external pressure.

Another reason for no effect may be that the control group was made aware of frequent SA and the risk of long-term SA by the invitational study leaflet and the Web-based questionnaire. The increased awareness may have stimulated the control group to take actions aimed at reducing their SA frequency; the process evaluation shows that they indeed did take actions.

Strengths and Limitations

This is the first RCT to study the effect of an EHI tool on employees with frequent SA. The baseline questionnaire used instruments and scales validated for use in occupational health care. As the SA measures were register-based, there was limited attrition. This study was undertaken in the Netherlands, limiting cross-country generalizability. The advantage is that all participants fall under the same regulations, with registration from the first SA day and salary payment also from the first SA day, regardless of the cause of SA.

As studies that include SA frequency as outcome measure are scarce and do not target the group of frequent absentees, we had no information on a realistic intervention effect. We included 2 sample size calculations in the study protocol as registered with the METc, 1 based on intervention studies in a more heterogeneous population and 1 based on our practice-based knowledge on a relevant intervention effect in an occupational health setting. Although aiming for the highest sample size, due to low participation, we accepted the lower sample size of 27 participants per group. In hindsight, this sample size was low and may be the reason for not finding effect.

Only 9.9% (82/825) eligible frequent absentees participated in the study. The absentees are probably not aware that frequent SA often recurs and may pose a risk of future long-term SA [11], regardless of the information in the invitational leaflet. We dealt with potential selection bias by using prestudy randomization: participants were allocated to intervention and control groups before the study started; they were blinded for the group to which they were allocated until they did or did not receive the digital advice. The low participation rate may have affected the generalizability of the results.

Few participants in the EHI-OP group visited the OP for advice and guidance. Thereupon, we conducted an analysis of the "combined intervention groups versus control group," which led to higher statistical power for testing the EHI than in the "analysis of the 3 study arms." However, we could not draw conclusions on the effectiveness of blended care (EHI-OP), as too few participants consulted the OP. The reason for low adherence to the blended intervention (EHI-OP) may be the voluntary character of the invitation, without any support or stimulation from researchers, employer, or OHS to visit the OP, in combination with an insufficient sense of urgency of the frequent absentee.

Learnings and Implications for Future Research

The study had a low participation rate (9.9%). This was probably due to a combination of factors. Employees with frequent SA may have a low intrinsic urgency to undertake action. The mild nature of most illnesses in frequent SA [50] in combination with an invitation related to sickness episodes in the past are possible reasons for a low intrinsic urgency to take action. In addition, employees are unaware of the high risk of future SA [11]. Moreover, there was no extrinsic stimulus: participation was completely voluntary. Most employees thought that the manager or colleagues did not find frequent SA a problem. A different recruitment procedure could increase study participation. A recent systematic review found that personal approaches lead to higher participation rates [51]. Recruitment can be increased through personal invitations by the researchers, invitation at the time of the third SA episode, when an employee has increased awareness possibly leading to increased motivation, or more involvement from the employer [52] or employee representatives. Due to scientific ethical codes (voluntary participation), involvement of the employer may be difficult to implement in a research setting. Several EHI studies found that recruitment using social media such as Facebook is more successful than more traditional recruitment strategies such as advertising [53,54]. However, such strategies seem difficult to apply to the population of employees with frequent SA. A combination of recruitment methods might also be more successful than relying solely on 1 method [55]. As it still may be difficult to reach participation rates as found in occupational health surveys (40% to 60%) and in earlier studies by Kant et al (50.2% questionnaire responders and 89% follow-up of [OP] intervention [15]) and Taimela et al (48% questionnaire responders and 68% attended the consultation at the OHS [14]), recruitment in intervention studies to reduce frequent SA should consider using a larger source population than the one used in this study. A higher participation rate combined with a larger source population may improve generalizability, decrease differences in characteristics of the participants between study arms, and increase power.

The adherence to the intervention tool was low. Reasons mentioned by participants were already having taken (a lot of) actions before the study or still undertaking actions, not seeing the added value of the intervention when knowing the problem not being work-related, being too busy, feeling low urgency, and being optimistic about their future health and SA. These reasons can be summarized as “low motivation to take (further) action.” To increase adherence, these issues have to be addressed. With little internal motivation, some external information or stimulus is needed. Studies on EHIs have found that human support increases adherence to the intervention [48,56] and effectiveness [49]. Increased adherence may be a more detailed explanation of the finding of Hutchesson in a meta-analysis [49] that direct human contact seems to intensify the effect of eHealth technologies. An OP can support various phases of the motivational process. An OP can increase readiness to act by increasing awareness of the high future risk of long-term SA. In addition, an OP can support an employee to make a realistic “plan of action” or refer to other professionals where necessary. An employer could also provide stimulation (they do in practice), but this is not ethically acceptable in scientific studies. Additional possibilities to increase adherence to the tool are the usage of reminders [57] or extra technological components such as more interactive exercises [58].

To increase adherence to an OP consultation, it might be more effective to offer only blended care, with the OP consultation as the main intervention and the EHI in the supporting role, as preparation for the consultation. Personal invitations to the OP could also increase adherence [51]. Preferably, the invitation is to the OP of the organization the employee works for; alternatively, a visit to another OP might be offered. The invitation should make clear that employees are also welcome in case of nonwork-related problems: the involvement of an occupational health service may be the reason why some participants in this study saw no reason for further action as they had private problems.

Future intervention studies on frequent absentees should deal with possible selection bias toward participants with more severe conditions by, for example, stratifying into groups with and without chronic disease or with and without long-term SA at baseline. Future intervention studies on SA frequency among employees with frequent SA should include larger sample sizes. Future research should also take into account the large reduction of SA frequency in the control group in the follow-up year. An intervention effect of 1 SA episode seems too high compared with the median SA frequency in the follow-up year. An intervention effect of 0.5 or even lower should be considered in future power calculations.

As the reasons for SA are very broad, it may be easier to develop an intervention focused on a specific disease than an intervention to reduce SA. However, it is important to continue to address reduction of SA in effectiveness studies. Not only is registered SA a very objective, numerical measurement that directly reflects economic costs (lost working days) but reduction of future frequent and long-term SA also focuses on prevention of any disease and ill-health.

Conclusions

To our knowledge, this is the first RCT to examine the effect of an EHI tool on employees with frequent SA. When comparing the intervention groups with the control group, we found no significant effect of the intervention on SA frequency. Moreover, we found no significant effect on total SA days, burnout, engagement, or work ability. Only few employees with frequent SA participated in the study and relatively few took additional action to reduce frequent SA, with no apparent difference between the intervention groups and the control group. None of the participants set up a “plan of action” aimed at reducing SA frequency. Due to low adherence to the OP consultation, the results on the EHI-OP study arm do not represent results from blended care. Future research should test the effect of blended care, possibly involving the employer or manager of the employee with frequent SA to encourage adherence to an OP consultation. This could help to increase awareness and sense of urgency and may lead to more focus and adequate (local professional) and appropriate support.

Acknowledgments

The authors would like to thank ArboNed for investing in the intervention as proposed. The authors would also like to thank Byelex, the software provider, who made the application to specification and enabled running the intervention. The authors AN, WvR, and CR worked for ArboNed. The authors have no support or funding to report, apart from the investment by ArboNed in the intervention. ArboNed had no influence on the design, analysis, results, or presentation of the study results.

Authors' Contributions

AN conceived the study. AN, CR, JG, and WvR designed the study. AN and CR performed the study. CR conducted the analyses, and AN wrote the first draft of the manuscript. All authors contributed to interpretation of the data, critical revision of the draft, and read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Informed consent form.

[PDF File (Adobe PDF File), 209KB - jmir_v20i10e10821_app1.pdf]

Multimedia Appendix 2

CONSORT - EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 2MB - jmir_v20i10e10821_app2.pdf]

Multimedia Appendix 3

Intervention elements.

[PDF File (Adobe PDF File), 27KB - jmir_v20i10e10821_app3.pdf]

Multimedia Appendix 4

Occupational physician instructions.

[PDF File (Adobe PDF File), 33KB - jmir_v20i10e10821_app4.pdf]

References

1. Virtanen M, Kivimäki M, Vahtera J, Elovainio M, Sund R, Virtanen P, et al. Sickness absence as a risk factor for job termination, unemployment, and disability pension among temporary and permanent employees. Occup Environ Med 2006 Mar;63(3):212-217 [FREE Full text] [doi: 10.1136/oem.2005.020297] [Medline: 16497865]
2. Monitorarbeid. De kosten van Ziekteverzuim voor werkgevers in Nederland URL: https://www.monitorarbeid.tno.nl/dynamics/modules/SPUB0102/view.php?pub_Id=100294&att_Id=4911 [accessed 2018-07-20] [WebCite Cache ID 713DdMYrY]
3. Koopmans PC, Roelen CA, Groothoff JW. Risk of future sickness absence in frequent and long-term absentees. Occup Med (Lond) 2008 Jun;58(4):268-274. [doi: 10.1093/occmed/kqn040] [Medline: 18390880]
4. Eakin J, MacEachen E. Health and the social relations of work: a study of the health-related experiences of employees in small workplaces. Sociol Health Ill 2001;20:894-914 [FREE Full text] [doi: 10.1111/1467-9566.00134]
5. Koopmans PC, Roelen CA, Groothoff JW. Frequent and long-term absence as a risk factor for work disability and job termination among employees in the private sector. Occup Environ Med 2008 Jul;65(7):494-499. [doi: 10.1136/oem.2007.034322] [Medline: 18032531]
6. Beemsterboer W, Stewart R, Groothoff J, Nijhuis F. A literature review on sick leave determinants (1984-2004). Int J Occup Med Environ Health 2009;22(2):169-179. [doi: 10.2478/v10001-009-0013-8] [Medline: 19617195]
7. Roelen CA, Schreuder JA, Koopmans PC, Moen BE, Groothoff JW. Sickness absence frequency among women working in hospital care. Occup Med (Lond) 2009 Oct;59(7):502-505. [doi: 10.1093/occmed/kqp089] [Medline: 19528330]
8. ten Brummelhuis LL, ter Hoeven CL, de Jong MD, Peper B. Exploring the linkage between the home domain and absence from work: health, motivation, or both? J Organiz Behav 2012 Feb 23;34(3):273-290. [doi: 10.1002/job.1789]
9. Roskes K, Donders NC, van der Gulden JW. Health-related and work-related aspects associated with sick leave: a comparison of chronically ill and non-chronically ill workers. Int Arch Occup Environ Health 2005 May;78(4):270-278. [doi: 10.1007/s00420-004-0596-0] [Medline: 15791473]
10. Notenbomer A, Groothoff JW, van Rhenen RW, Roelen CA. Associations of work ability with frequent and long-term sickness absence. Occup Med (Lond) 2015 Jul;65(5):373-379. [doi: 10.1093/occmed/kqv052] [Medline: 25964509]
11. Notenbomer A, Roelen CA, van Rhenen W, Groothoff JW. Focus group study exploring factors related to frequent sickness absence. PLoS One 2016;11(2):e0148647 [FREE Full text] [doi: 10.1371/journal.pone.0148647] [Medline: 26872050]
12. Arends I, van der Klink JJ, van Rhenen W, de Boer MR, Bültmann U. Prevention of recurrent sickness absence in workers with common mental disorders: results of a cluster-randomised controlled trial. Occup Environ Med 2014 Jan;71(1):21-29. [doi: 10.1136/oemed-2013-101412] [Medline: 24158311]
13. Duijts SF, Kant I, van den Brandt PA, Swaen GM. Effectiveness of a preventive coaching intervention for employees at risk for sickness absence due to psychosocial health complaints: results of a randomized controlled trial. J Occup Environ Med 2008 Jul;50(7):765-776. [doi: 10.1097/JOM.0b013e3181651584] [Medline: 18617832]
14. Taimela S, Malmivaara A, Justén S, Läärä E, Sintonen H, Tiekso J, et al. The effectiveness of two occupational health intervention programmes in reducing sickness absence among employees at risk. Two randomised controlled trials. Occup Environ Med 2008 Apr;65(4):236-241 [FREE Full text] [doi: 10.1136/oem.2007.032706] [Medline: 17681994]
15. Kant I, Jansen NW, van Amelsvoort LG, van Leusden R, Berkouwer A. Structured early consultation with the occupational physician reduces sickness absence among office workers at high risk for long-term sickness absence: a randomized controlled trial. J Occup Rehabil 2008 Mar;18(1):79-86 [FREE Full text] [doi: 10.1007/s10926-007-9114-z] [Medline: 18196446]

16. Bolier L, Haverman M, Kramer J, Westerhof GJ, Riper H, Walburg JA, et al. An Internet-based intervention to promote mental fitness for mildly depressed adults: randomized controlled trial. J Med Internet Res 2013;15(9):e200 [FREE Full text] [doi: 10.2196/jmir.2603] [Medline: 24041479]
17. Newhouse N, Martin A, Jawad S, Yu L, Davoudianfar M, Locock L, et al. Randomised feasibility study of a novel experience-based internet intervention to support self-management in chronic asthma. BMJ Open 2016 Dec 28;6(12):e013401 [FREE Full text] [doi: 10.1136/bmjopen-2016-013401] [Medline: 28031210]
18. van Genugten L, van Empelen P, Boon B, Borsboom G, Visscher T, Oenema A. Results from an online computer-tailored weight management intervention for overweight adults: randomized controlled trial. J Med Internet Res 2012;14(2):e44 [FREE Full text] [doi: 10.2196/jmir.1901] [Medline: 22417813]
19. van Beugen S, Ferwerda M, Hoeve D, Rovers MM, Spillekom-van KS, 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]
20. Beishuizen CR, Stephan BC, van Gool WA, Brayne C, Peters RJ, Andrieu S, et al. Web-based interventions targeting cardiovascular risk factors in middle-aged and older people: a systematic review and meta-analysis. J Med Internet Res 2016 Mar 11;18(3):e55 [FREE Full text] [doi: 10.2196/jmir.5218] [Medline: 26968879]
21. Riper H, Spek V, Boon B, Conijn B, Kramer J, Martin-Abello K, et al. Effectiveness of E-self-help interventions for curbing adult problem drinking: a meta-analysis. J Med Internet Res 2011;13(2):e42 [FREE Full text] [doi: 10.2196/jmir.1691] [Medline: 21719411]
22. Beiwinkel T, Eißing T, Telle N, Siegmund-Schultze E, Rössler W. Effectiveness of a web-based intervention in reducing depression and sickness absence: randomized controlled trial. J Med Internet Res 2017 Jun 15;19(6):e213 [FREE Full text] [doi: 10.2196/jmir.6546] [Medline: 28619701]
23. Ebert DD, Lehr D, Boß L, Riper H, Cuijpers P, Andersson G, et al. Efficacy of an internet-based problem-solving training for teachers: results of a randomized controlled trial. Scand J Work Environ Health 2014 Nov;40(6):582-596. [doi: 10.5271/sjweh.3449] [Medline: 25121986]
24. Ebert DD, Heber E, Berking M, Riper H, Cuijpers P, Funk B, et al. Self-guided internet-based and mobile-based stress management for employees: results of a randomised controlled trial. Occup Environ Med 2016 May;73(5):315-323. [doi: 10.1136/oemed-2015-103269] [Medline: 26884049]
25. Ebert DD, Lehr D, Heber E, Riper H, Cuijpers P, Berking M. Internet- and mobile-based stress management for employees with adherence-focused guidance: efficacy and mechanism of change. Scand J Work Environ Health 2016 Jun 01;42(5):382-394. [doi: 10.5271/sjweh.3573]
26. Heber E, Lehr D, Ebert DD, Berking M, Riper H. Web-based and mobile stress management intervention for employees: results of a randomised controlled trial. J Med Internet Res 2016 Jan 27;18(1):e21 [FREE Full text] [doi: 10.2196/jmir.5112] [Medline: 26818683]
27. Volker D, Zijlstra-Vlasveld MC, Anema JR, Beekman AT, Brouwers EP, Emons WH, et al. Effectiveness of a blended web-based intervention on return to work for sick-listed employees with common mental disorders: results of a cluster randomized controlled trial. J Med Internet Res 2015;17(5):e116 [FREE Full text] [doi: 10.2196/jmir.4097] [Medline: 25972279]
28. Demerouti E, Bakker AB, Nachreiner F, Schaufeli WB. The job demands-resources model of burnout. J Appl Psychol 2001 Jun;86(3):499-512. [Medline: 11419809]
29. Clausen T, Burr H, Borg V. Do psychosocial job demands and job resources predict long-term sickness absence? An analysis of register-based outcomes using pooled data on 39,408 individuals in four occupational groups. Int Arch Occup Environ Health 2014 Nov;87(8):909-917. [doi: 10.1007/s00420-014-0936-7] [Medline: 24562968]
30. Slany C, Schütte S, Chastang J, Parent-Thirion A, Vermeylen G, Niedhammer I. Psychosocial work factors and long sickness absence in Europe. Int J Occup Environ Health 2014;20(1):16-25 [FREE Full text] [doi: 10.1179/2049396713Y.0000000048] [Medline: 24176393]
31. Borritz M, Christensen KB, Bültmann U, Rugulies R, Lund T, Andersen I, et al. Impact of burnout and psychosocial work characteristics on future long-term sickness absence. Prospective results of the Danish PUMA Study among human service workers. J Occup Environ Med 2010 Oct;52(10):964-970. [doi: 10.1097/JOM.0b013e3181f12f95] [Medline: 20881631]
32. Schaufeli WB, Bakker AB, Van Rhenen W. How changes in job demands and resources predict burnout, work engagement, and sickness absenteeism. J Organiz Behav 2009 Oct;30(7):893-917. [doi: 10.1002/job.595]
33. nvab-online. Occupational physician guideline for the treatment of workers with mental health problems URL: https://www.nvab-online.nl/sites/default/files/bestanden-webpaginas/MDRL_Overspanning-Burnout.pdf [accessed 2018-09-28] [WebCite Cache ID 72lmdjZUB]
34. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
35. Wet verbetering Poortwachter. URL: http://wetten.overheid.nl/BWBR0013063/2008-11-01 [WebCite Cache ID 72lnJmtUV]
36. van Dierendonck S. Handleiding van de Utrechtse Burnout Schaal (UBOS) (Manual Utrecht Burnout Scale). Lisse, the Netherlands: Swets Test Services; 2000.

37. Schaufeli WB. The measurement of work engagement with a short questionnaire: a cross-national study. Educ Psychol Meas 2006 Aug 01;66(4):701-716. [doi: 10.1177/0013164405282471]
38. Jääskeläinen A, Kausto J, Seitsamo J, Ojajärvi A, Nygård C, Arjas E, et al. Work ability index and perceived work ability as predictors of disability pension: a prospective study among Finnish municipal employees. Scand J Work Environ Health 2016 Dec 01;42(6):490-499. [doi: 10.5271/sjweh.3598] [Medline: 27706492]
39. El Fassi M, Bocquet V, Majery N, Lair ML, Couffignal S, Mairiaux P. Work ability assessment in a worker population: comparison and determinants of Work Ability Index and Work Ability score. BMC Public Health 2013 Apr 08;13:305 [FREE Full text] [doi: 10.1186/1471-2458-13-305] [Medline: 23565883]
40. Van Veldhoven M, Meijman T. Het meten van psychosociale arbeidsbelasting met een vragenlijst. Amsterdam, the Netherlands: NIA; 1994.
41. Halford C, Wallman T, Welin L, Rosengren A, Bardel A, Johansson S, et al. Effects of self-rated health on sick leave, disability pension, hospital admissions and mortality. A population-based longitudinal study of nearly 15,000 observations among Swedish women and men. BMC Public Health 2012 Dec 22;12:1103 [FREE Full text] [doi: 10.1186/1471-2458-12-1103] [Medline: 23259777]
42. DeSalvo KB, Fan VS, McDonell MB, Fihn SD. Predicting mortality and healthcare utilization with a single question. Health Serv Res 2005 Aug;40(4):1234-1246 [FREE Full text] [doi: 10.1111/j.1475-6773.2005.00404.x] [Medline: 16033502]
43. 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]
44. Kemper H, Ooijendijk W, Stiggelbout M. Consensus over de nederlandse norm voor gezond bewegen. 2000. URL: https://repository.tudelft.nl/view/tno/uuid:a8cabc4a-e5d2-422e-aea9-449becabd5a4 [WebCite Cache ID 72lozW8BE]
45. Douwes M, Hildebrandt V. Vragen naar de mate van lichamelijke activiteit; onderzoek naar de test-hertest betrouwbaarheid en congruente validiteit van een vragenlijst. 2000. URL: https://repository.tudelft.nl/view/tno/uuid%3Aab5060df-f1fd-4b03-8d65-ccba1c94f005 [WebCite Cache ID 734LwBRge]
46. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [Medline: 17695343]
47. Random Integer Generator. URL: https://www.random.org/integers/ [WebCite Cache ID 736fCGakm]
48. Heber E, Ebert DD, Lehr D, Cuijpers P, Berking M, Nobis S, et al. The benefit of web- and computer-based interventions for stress: a systematic review and meta-analysis. J Med Internet Res 2017 Feb 17;19(2):e32 [FREE Full text] [doi: 10.2196/jmir.5774] [Medline: 28213341]
49. Hutchesson MJ, Rollo ME, Krukowski R, Ells L, Harvey J, Morgan PJ, et al. eHealth interventions for the prevention and treatment of overweight and obesity in adults: a systematic review with meta-analysis. Obes Rev 2015 May;16(5):376-392. [doi: 10.1111/obr.12268] [Medline: 25753009]
50. Hörnquist JO, Hansson B, Leijon M, Mikaelsson B. Repeated short-term sick-leave and quality of life. An evaluation of a clinical socio-medical intervention. Scand J Soc Med 1990 Jun;18(2):91-95. [Medline: 2367826]
51. van Zon SK, Scholtens S, Reijneveld SA, Smidt N, Bültmann U. Active recruitment and limited participant-load related to high participation in large population-based biobank studies. J Clin Epidemiol 2016 Oct;78:52-62 [FREE Full text] [doi: 10.1016/j.jclinepi.2016.03.009] [Medline: 27032874]
52. Arends I, Bültmann U, Shaw WS, van Rhenen W, Roelen C, Nielsen K, et al. How to engage occupational physicians in recruitment of research participants: a mixed-methods study of challenges and opportunities. J Occup Rehabil 2014 Mar;24(1):68-78. [doi: 10.1007/s10926-013-9452-y] [Medline: 23690088]
53. Leonard A, Hutchesson M, Patterson A, Chalmers K, Collins C. Recruitment and retention of young women into nutrition research studies: practical considerations. Trials 2014;15:23 [FREE Full text] [doi: 10.1186/1745-6215-15-23] [Medline: 24433229]
54. Watson NL, Mull KE, Heffner JL, McClure JB, Bricker JB. Participant recruitment and retention in remote eHealth intervention trials: methods and lessons learned from a large randomized controlled trial of two web-based smoking interventions. J Med Internet Res 2018 Aug 24;20(8):e10351 [FREE Full text] [doi: 10.2196/10351] [Medline: 30143479]
55. Hou S, Charlery SR, Roberson K. Systematic literature review of internet interventions across health behaviors. Health Psychol Behav Med 2014 Jan 01;2(1):455-481 [FREE Full text] [doi: 10.1080/21642850.2014.895368] [Medline: 25750795]
56. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
57. Fry JP, Neff RA. Periodic prompts and reminders in health promotion and health behavior interventions: systematic review. J Med Internet Res 2009;11(2):e16 [FREE Full text] [doi: 10.2196/jmir.1138] [Medline: 19632970]
58. Hasson H, Brown C, Hasson D. Factors associated with high use of a workplace web-based stress management program in a randomized controlled intervention study. Health Educ Res 2010 Aug;25(4):596-607 [FREE Full text] [doi: 10.1093/her/cyq005] [Medline: 20150531]

Abbreviations

CAU: care as usual
eHealth: electronic health
EHI: eHealth intervention
EHI-only: study arm with only eHealth intervention
EHI-OP: study arm with eHealth intervention and invitation for occupational physician consultation
GP: general practitioner
IQR: interquartile range
METc: medical ethics committee
NVAB: Netherlands Society for Occupational Physicians
OP: occupational physician
OHS: occupational health service
RCT: randomized controlled trial
SA: sickness absence
UBOS: 9-item Utrecht Burnout Scale
UWES: 9-item Utrecht Work Engagement Scale
WAI: Work Ability Index

Edited by G Eysenbach; submitted 20.04.18; peer-reviewed by D Ebert, E Siegmund-Schultze, L Boß; comments to author 09.06.18; revised version received 20.07.18; accepted 25.09.18; published 23.10.18.

Please cite as:
Notenbomer A, Roelen C, Groothoff J, van Rhenen W, Bültmann U
Effect of an eHealth Intervention to Reduce Sickness Absence Frequency Among Employees With Frequent Sickness Absence: Randomized Controlled Trial
J Med Internet Res 2018;20(10):e10821
URL: http://www.jmir.org/2018/10/e10821/
doi:10.2196/10821
PMID:30355551

Original Paper

Geographic Representativeness of a Web-Based Smoking Cessation Intervention: Reach Equity Analysis

Michael S Amato[1], PhD; Amanda L Graham[1,2], PhD

[1]Schroeder Institute, Truth Initiative, Washington, DC, United States

[2]Department of Oncology, Georgetown University Medical Center / Cancer Prevention and Control Program, Lombardi Comprehensive Cancer Center, Washington, DC, United States

Corresponding Author:
Michael S Amato, PhD
Schroeder Institute
Truth Initiative
900 G Street NW
Fourth Floor
Washington, DC, 20001
United States
Phone: 1 202 454 5767
Fax: 1 202 454 5555
Email: mamato@truthinitiative.org

Abstract

Background: Cigarette smoking is the leading cause of preventable death and disease in the United States. Smoking prevalence is higher in rural areas than in metropolitan areas, due partly to differences in access to cessation treatment. With internet use at 89% of all US adults, digital approaches could increase use of cessation treatment and reduce smoking.

Objective: We investigated the extent to which smokers from rural areas use a digital cessation resource. We compared the geographic distribution of registered users of a free Web-based smoking cessation program with the geographic distribution of US smokers.

Methods: We mapped user-provided ZIP codes to Rural-Urban Continuum Codes. A total of 59,050 of 118,574 users (49.80%) provided valid ZIP codes from 2013 to 2017. We used US National Survey of Drug Use and Health data from 2013 to 2017 to compare the geographic distribution of our sample of Web-based cessation users with the geographic distribution of US smokers. Reach ratios and 95% confidence intervals quantified the extent to which rural smokers' representation in the sample was proportionate to their representation in the national smoking population. Reach ratios less than 1 indicate underrepresentation.

Results: Smokers from rural areas were significantly underrepresented in 2013 (reach ratio 0.89, 95% CI 0.87-0.91) and 2014 (reach ratio 0.89, 95% CI 0.86-0.92), proportionally represented in 2015 (reach ratio 1.08, 95% CI 1.02-1.14) and 2016 (reach ratio 1.03, 95% CI 0.94-1.14), and proportionally overrepresented in 2017 (reach ratio 1.16, 95% CI 1.12-1.21). Smokers from Large Metro areas were proportionally represented in 2013 and 2014 but underrepresented in 2015 (reach ratio 0.97, 95% CI 0.94-1.00), 2016 (reach ratio 0.89, 95% CI 0.85-0.94), and 2017 (reach ratio 0.89, 95% CI 0.86-0.91).

Conclusions: Results suggest that smokers from rural areas are more than proportionally reached by a long-standing digital cessation intervention. The underrepresentation of smokers from Large Metro areas warrants further study.

(J Med Internet Res 2018;20(10):e11668) doi:10.2196/11668

KEYWORDS

smoking cessation; health behavior; internet; population health; rural health; urban health; health equity; telemedicine

Introduction

Geographic Disparities in Smoking Prevalence

The disease burden from cigarette smoking—still the leading cause of preventable death and disease in the United States [1]—disproportionately affects rural Americans [2-4]. Although the national prevalence of cigarette smoking was 13.9% in 2017, sharp geographic disparities exist: adults living outside metropolitan areas were nearly twice as likely to smoke as their urban counterparts (21.5% and 11.5%, respectively) [5]. While there has been a substantial decline in smoking prevalence in urban populations, smoking rates between 2007 and 2014 remained stagnant in rural areas [6]. Rural residence has been shown to be an independent predictor of this difference in smoking trends, even after controlling for important covariates [6-8]. Understanding this geographic disparity and reducing the higher rates of smoking among rural Americans has been acknowledged as a public health priority [9,10].

Rural populations and cultures are heterogeneous, as are the challenges they face when quitting tobacco [7,11]. Among these are barriers to treatment access, including lack of insurance [12], geographic isolation, and limited access to trained tobacco treatment providers [10]. The internet is often the first place that many smokers turn to for health information and is well suited to address barriers to cessation treatment access for rural adults. In 2017, 36% of all smokers in the United States—more than 12 million adults—searched online for information about quitting smoking [13]. Evidence-based internet interventions deliver the core components of cessation treatment through engaging, multimodal formats, often free of charge to the end user, and yield quit rates comparable with in-person and phone-based interventions [14,15]. They are available around the clock and (uniquely) can be accessed conveniently at times when smokers most need support to prevent relapse. Thriving, open access online social networks for smoking cessation (eg, BecomeAnEX [16]) provide connection to a broad range of current and former smokers for real-time information and support that may be lacking in rural smokers' own social networks [17].

However, enthusiasm for the potential of internet interventions to reduce smoking prevalence among rural adults may need to be tempered by the realities of a persistent digital divide. According to the Pew Research Center, 89% of all American adults used the internet in 2018, yet disparities in usage rates remain [18]. For almost two decades, rural adults have been roughly 10 percentage points less likely to use the internet than their urban and suburban counterparts. In 2018, 78% of rural adults reported that they used the internet, compared with 92% of urban adults and 90% of suburban adults [18]. In addition to differences in access, differences in bandwidth may affect rural adults' ability to take advantage of rich, interactive features: as of 2016, broadband was available to 96% of urban Americans but only 61% of rural Americans [19].

To our knowledge, only 1 study has examined the reach of internet smoking cessation interventions across the rural-urban continuum. In 2007, Danaher et al [20] compared the proportion of rural adults enrolled in a Web-based trial for smokeless tobacco cessation with the proportion of rural adults in the US national population. Using a chi-square test, they found that a significantly greater proportion of trial participants lived in rural areas (8.1%) than would be expected based on the national proportion of adults who lived in rural areas (4.2%). That result reflected both smokeless tobacco usage patterns and the success of the study's targeted marketing efforts.

Objectives

In this study, we extended the work of Danaher et al to investigate the participation of rural US smokers in a free, evidence-based, digital smoking cessation intervention. Instead of the chi-square statistic, we employed the reach ratio (ReRa) [21], a measure of population reach that has been used to document the extent to which specific subgroups of a population benefit from a public health intervention [21-23]. The benefit of the ReRa method is that it provides not only a significance test, but also an estimate of magnitude and confidence interval for numeric comparison. A ReRa of 1.0 indicates that a group is perfectly represented, greater than 1.0 indicates overrepresentation, and less than 1.0 indicates underrepresentation. For example, Baskerville et al [22] used ReRas to examine the equity of quitline treatment reach across 3 specific vulnerable populations (young males; those with less than high school education; and rural dwellers) following implementation of tobacco warning labels with the toll-free quitline number. They found substantial variability in the reach equity of quitlines for rural populations across Canadian provinces, ranging from approximately 0.1 for rural smokers in Manitoba to approximately 1.4 for rural smokers in New Brunswick. Similarly, Amato et al [23] examined quitline use in Minnesota and found that rural Minnesotans were less likely than their urban counterparts to use quitline services (ReRa 0.78). In this study, we used ReRas to measure the extent to which the proportion of smokers who registered for a free Web-based smoking cessation program from rural areas matched their representation in the national population.

Methods

Study Setting

We extracted data from 2013 to 2017 from BecomeAnEX, a free Web-based smoking cessation program developed by Truth Initiative (Washington, DC, USA) in collaboration with Mayo Clinic (Rochester, MN, USA). The program was launched in 2008 [24] and delivers cessation treatment consistent with US national treatment guidelines [25]. Smokers find BecomeAnEX through organic search, using common search phrases like "how to quit smoking." Smokers also find BecomeAnEX through paid advertisements on Google search. Smokers anywhere in the United States had equal opportunity to register on BecomeAnEX (ie, no geotargeting of advertisements was in place during the study period). Upon registration, users provide an email address, choose a username, and designate a password. During the study period, users could also choose to provide optional pieces of personal information at registration, including ZIP code, age, gender, and smoking history.

Procedures for collecting ZIP code information on BecomeAnEX changed during the study period. During most years, ZIP code was a required element of the registration process; during 2015, 2016, and part of 2017, it was entered optionally. As a result, the proportions of users who entered a ZIP code during those years were substantially lower than during others. To ensure that estimates of reach were not driven by sample differences, we assessed the effects of gender and age on likelihood of reporting a valid ZIP code using separate quasi-Poisson logistic regressions. Age was entered as a continuous variable measured in decades; female gender was entered as an indicator variable. Interaction terms with year were entered as categoric variables.

Geographic Classification of Users

We used ZIP codes from user registration data to calculate ReRas following 2 preprocessing steps. In the first preprocessing step, we classified BecomeAnEX users in terms of location. We mapped each user's ZIP code to a 2013 US Department of Agriculture Rural-Urban Continuum Codes (RUCC) value. RUCC provides a classification system for distinguishing metropolitan counties based on the size of the metropolitan area, and for nonmetropolitan counties based on the extent of urbanization and proximity to metropolitan areas, drawing on the 2010 US Census and the 2006-10 American Community Survey [26]. We mapped ZIP codes to ZIP Code Tabulation Areas (ZCTAs), then ZCTA to county, and finally county to 2013 RUCC [27]. Demographic differences in the likelihood of providing a ZIP code during registration were assessed with logistic regression.

We selected the RUCC classification system to allow direct comparison with detailed tables from the US National Survey of Drug Use and Health (NSDUH), which also report geographic area using the RUCC system. Although RUCC uses a 9-point classification system, we combined geographic area types into the following 3 categories to ensure that samples in all groups were of sufficient size to support meaningful inference based on confidence intervals: Large Metro (RUCC 1), Small Metro (RUCC 2 or 3), and Nonmetro (RUCC 4-9) [27]. Large Metro areas had populations of 1 million or more people. Small Metro areas had populations of fewer than 1 million people but were defined as metropolitan areas based on population and worker commuting criteria by the US Office of Management and Budget. Briefly, those criteria included an urbanized core of 50,000 or more population and adjacent areas that were socially and economically integrated [28]. Nonmetro areas were counties outside of defined metropolitan areas [26]. Although our definition of Nonmetro was not strictly limited to individuals living in areas classified as "completely rural" (RUCC 9), it included the roughly 20% of US smokers living at the rural end of the rural-urban continuum. Previous research has also grouped geographic classifications [2-4,6,7,10].

Calculation of Reach Ratios

In a second preprocessing step, we calculated numerators (proportions of BecomeAnEX users) and denominators (proportions of US smokers) for the ReRas. We defined numerators for ReRas as the proportions of BecomeAnEX users from each geographic category within each year. These proportions were calculated from the geotagged BecomeAnEX registration data. For example, in 2016, 43.8% of new users on BecomeAnEX lived in Large Metro areas (Table 1, bottom panel).

We defined denominators for ReRas as the proportions of US smokers from rural and urban areas. These proportions were calculated from NSDUH estimates of the numbers of smokers by geographic category (Table 2.56A for 2013 [29]; Table 2.56A for 2014 [30]; Table 2.41A for 2015 and 2016 [31]; and Table 2.41A in 2017 [32]). Proportions were calculated as the estimated number of smokers within each geographic category, divided by the total number of smokers. For example, in 2016 there were 51,333,000 smokers aged 12 years and over living in the United States, of whom 25,259,000 lived in Large Metro areas (49.2%; Table 1, top panel; Table 2.41 in 2016 NSDUH detailed tables [31]).

Finally, we created ReRas from the 2 sets of proportions. Following the approach of Campbell et al [21], we calculated 95% confidence intervals for each ReRa using the Wald interval method (section 4.2.1 in Fagerland et al [33]). The key question of interest was whether ReRas for each group significantly differed from 1.0 in any years.

Table 1. Proportions of smokers in the US national population and among users of a Web-based intervention, by location and year.

Population and type of area (RUCC[a])	Year, %				
	2013	2014	2015	2016	2017
US smokers[b]					
Large Metro (1)	48.14	50.07	50.29	49.21	48.18
Small Metro (2,3)	31.63	31.00	32.93	31.99	33.65
Nonmetro (4-9)	20.23	18.93	16.78	18.80	18.17
Geographically classified BecomeAnEX users[c]					
Large Metro (1)	48.86	50.63	48.86	43.85	42.65
Small Metro (2,3)	33.16	32.53	33.05	36.70	36.24
Nonmetro (4-9)	17.98	16.83	18.08	19.45	21.11

[a] Rural-Urban Continuum Codes.

[b] Total numbers of smokers (in thousands) aged ≥12 years living in the United States were 55,778 in 2013; 55,240 in 2014; 51,951 in 2015; 51,333 in 2016; and 48,692 in 2017.

[c] Total numbers of BecomeAnEX users who reported a valid ZIP code were 33,484 in 2013; 18,255 in 2014; 5491 in 2015; 1820 in 2016; and 8832 in 2017.

Results

Sample Characteristics

Between 2013 and 2017, a total of 127,207 new users registered on BecomeAnEX. Among those users, 67,854 provided valid ZIP codes and were assigned a RUCC value, representing 53.34% of all new users during the time period. Virtually all (114,373/114,844, 99.95%) users were current smokers or former smokers. Age was provided by 77.22% (98,227/127,207) of users; gender was provided by 76.61% (97,449/127,207). For every additional decade of a user's age at registration, they were on average 3% more likely to provide a ZIP code across all years in the study period (relative risk 1.03, 95% CI 1.027-1.031, ranging from 1.00 in 2016 to 1.03 in 2013). Women were 10% more likely than men to provide a ZIP code (relative risk 1.10, 95% CI 1.103-1.12, ranging from 1.01 in 2015 to 1.44 in 2016).

Table 1 shows the proportions of smokers in each geographic area by year, for both the national population and the treatment-seeking population of BecomeAnEX users. In both samples, the largest group in all years lived in Large Metro areas (43.85% to 50.63%), while the smallest group lived in Nonmetro areas (16.83% to 21.11%). About a third in both samples lived in Small Metro areas (31.00% to 36.70%).

Table 2 shows the extent to which the geographic distribution of smokers seeking treatment on BecomeAnEX mirrored the geographic distribution of smokers in the general population, with ReRas and 95% confidence intervals.

Large Metro Areas

In 2013 and 2014, the proportions of BecomeAnEX users from Large Metro areas (RUCC=1) were similar to their representation in the general smoking population. Beginning in 2015, smokers from Large Metro areas were marginally underrepresented (ReRa 0.97, 95% CI 0.94-1.00), and in 2016 and 2017 they were underrepresented by an even greater extent (2016: ReRa 0.89, 95% CI 0.85-0.94; 2017: ReRa 0.89, 95% CI 0.86-0.91).

Small Metro Areas

In contrast, smokers from Small Metro areas (RUCC 2 or 3) were overrepresented in all years except 2015. The magnitude of the disparity was greatest in 2016 (2013: ReRa 1.05, 95% CI 1.03-10.7; 2014: ReRa 1.05, 95% CI 1.02-1.08; 2016: ReRa 1.15, 95% CI 1.08-1.22; 2017: ReRa 1.08, 95% CI 1.05-1.11).

Nonmetro Areas

The representation of Nonmetro smokers (RUCC 4-9) on BecomeAnEX increased the most during the study period. They were underrepresented in 2013 (ReRa 0.89, 95% CI 0.87-0.91) and 2014 (ReRa 0.89, 95% CI 0.86-0.92), overrepresented in 2015 (ReRa 1.08, 95% CI 1.02-1.14), proportionally represented in 2016 (ReRa 1.03, 95% CI 0.94-1.14), and then overrepresented again in 2017 (ReRa 1.16, 95% CI 1.12-1.21).

Table 2. Reach ratios (ReRas) and 95% confidence intervals for the geographic distribution of a Web-based smoking cessation intervention in the United States.

Type of area (RUCC[a])	Year, ReRa (95% CI)				
	2013	2014	2015	2016	2017
Large Metro (1)	1.01 (1.00-1.03)	1.01 (0.99-1.03)	0.97 (0.94-1.00)	0.89 (0.85-0.94)[b]	0.89 (0.86-0.91)[b]
Small Metro (2,3)	1.05 (1.03-1.07)[b]	1.05 (1.02-1.08)[b]	1.00 (0.96-1.04)[b]	1.15 (1.08-1.22)[b]	1.08 (1.05-1.11)[b]
Nonmetro (4-9)	0.89 (0.87-0.91)[b]	0.89 (0.86-0.92)[b]	1.08 (1.02-1.14)[b]	1.03 (0.94-1.14)	1.16 (1.12-1.21)[b]

[a] Rural-Urban Continuum Codes.

[b] ReRa significantly different from 1.

Discussion

Principal Results

This study examined the extent to which registered users on a free Web-based smoking cessation program were geographically representative of the US national smoking population. Results indicate that the reach of BecomeAnEX was relatively similar across the rural-urban continuum. We observed some differences, although the magnitude of difference was relatively small when compared with other published ReRas [21,22]. The reach to rural areas was of particular interest, because smokers in those areas often have higher smoking rates and reduced access to other forms of cessation treatment.

Summarizing across all years, there is evidence for two general trends. First, despite the digital divide, smokers from Nonmetro areas appeared to seek out and use digital cessation resources at a rate that was proportional to—and even higher than—their representation in the population. The expansion of broadband availability and smartphone penetration may have facilitated the proportional increase from 2013 and 2014. Second, we observed a proportional decrease in BecomeAnEX users from Large Metro areas. One speculative hypothesis for that trend could be higher prevalence of nondaily and social smoking among smokers in Large Metro areas; many nondaily or social smokers do not view themselves as smokers and do not seek cessation treatment [34,35]. Alternatively, smokers in Large Metro areas might be using other types of cessation resources. Given the large proportion of the population living in Large Metro areas and the substantial population health benefits of quitting smoking, future research should explore this unexpected finding.

Limitations

These analyses have several limitations. First, the geographic classification of BecomeAnEX users and NSDUH respondents was based on 2013 RUCC codes, which are updated every 10 years following each US Census. The census is the best available data for our purpose but does not capture changes that may have occurred between individual years in the study period. Second, other rural-urban classification systems with more nuanced categories may have provided additional insight [36]. We chose to use the RUCC system because it is widely used throughout the literature. We chose to combine all Nonmetro RUCC classifications into a single category due to sample size limitations. It is possible that smokers in the least connected areas may face unique challenges to quitting that were not detected in our analysis. However, analyzing them independently resulted in unacceptably large confidence intervals that prevented meaningful interpretation (not reported). Future research should investigate how smokers in the most rural communities access evidence-based cessation treatment, using alternative methods. Nonetheless, the reported ReRas for the combined category provides a useful and previously unavailable estimate of treatment reach for 20% of the smoking population. Third, although the method used to create confidence intervals has been previously published and is the best available to our knowledge, it ignores variability in NSDUH smoking prevalence estimates, while simultaneously overestimating error variance in the ReRas for each individual year by ignoring cross-year correlations. Analytic techniques that overcome these methodological limitations are needed, such as a method for multinomial trend analysis of ReRas.

Fourth, ZIP codes were available for only 53.34% of registered users. Although older users and women were more likely than younger users and men to provide ZIP codes, we have no reason to expect systematic differences by age or gender between rural and urban US adults in their proclivity to provide ZIP codes during the registration process but cannot rule out the possibility. Differences in rates of ZIP code missingness would not be expected to affect the robustness of our findings, unless there were an interactive effect of gender or age with geographic area type on a smoker’s likelihood of seeking treatment from an internet intervention. We are not aware of any data to suggest such an interaction exists. Previous research has found that the geographic disparity in smoking rates is greater for women than for men [8]. That finding does not suggest disparities in treatment seeking and does not directly affect our conclusions, but warrants further study. Future research, using appropriate population surveillance methods that can control for multiple covariates, should further investigate demographic correlates of geographic disparities.

Finally, our research focused on one specific Web-based cessation program, and therefore our results are not synonymous with the reach of all digital cessation interventions across the rural-urban continuum. However, our approach provides an easily replicable model for other intervention platforms to document their reach based on geography.

Conclusions

Progress is needed in addressing other challenges facing rural communities. Comprehensive smoke-free air laws, changes in social norms around tobacco use, effective tobacco industry countermarketing, and limiting children’s access to tobacco products are among the changes that still need to reach many rural communities. However, with the broad (and increasing) reach and proven effectiveness of internet smoking cessation interventions, access to treatment should not remain a barrier to quitting among rural smokers.

Acknowledgments

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

Conflicts of Interest

Both authors are employed by Truth Initiative, which runs BecomeAnEX and This Is Quitting, digital cessation resources available free to consumers and licensed to tobacco control clients.

References

1. National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health. The health consequences of smoking—50 years of progress: a report of the Surgeon General. Atlanta, GA: Centers for Disease Control and Prevention; 2014.
2. Doescher MP, Jackson JE, Jerant A, Gary HL. Prevalence and trends in smoking: a national rural study. J Rural Health 2006;22(2):112-118. [doi: 10.1111/j.1748-0361.2006.00018.x] [Medline: 16606421]
3. Eberhardt M, Ingram D, Makuc D. Health, United States: Urban and rural health chartbook. Hyattsville, MD: National Center for Health Statistics; 2001. URL: https://www.cdc.gov/nchs/data/hus/hus01cht.pdf [accessed 2018-07-23] [WebCite Cache ID 7185H4jYH]
4. Tobacco Use by Geographic Region. Atlanta, GA: Office on Smoking, Centers for Disease Control and Prevention; 2018. URL: https://www.cdc.gov/tobacco/disparities/geographic/ [accessed 2018-07-23] [WebCite Cache ID 7185T44gz]
5. Norris T, Schiller J, Clarke TC. Early release of selected estimates based on data from the National Health Interview Survey. Atlanta, GA: National Center for Health Statistics, Centers for Disease Control and Prevention; 2018. URL: https://www.cdc.gov/nchs/data/nhis/earlyrelease/earlyrelease201705.pdf [accessed 2018-07-23] [WebCite Cache ID 718B5Z4WZ]

6. Doogan NJ, Roberts ME, Wewers ME, Stanton CA, Keith DR, Gaalema DE, et al. A growing geographic disparity: rural and urban cigarette smoking trends in the United States. Prev Med 2017 Nov;104:79-85. [doi: 10.1016/j.ypmed.2017.03.011] [Medline: 28315761]
7. Roberts ME, Doogan NJ, Kurti AN, Redner R, Gaalema DE, Stanton CA, et al. Rural tobacco use across the United States: how rural and urban areas differ, broken down by census regions and divisions. Health Place 2016 Dec;39:153-159 [FREE Full text] [doi: 10.1016/j.healthplace.2016.04.001] [Medline: 27107746]
8. Cepeda-Benito A, Doogan NJ, Redner R, Roberts ME, Kurti AN, Villanti AC, et al. Trend differences in men and women in rural and urban U.S. settings. Prev Med 2018 Apr 05:9-20. [doi: 10.1016/j.ypmed.2018.04.008] [Medline: 29627511]
9. US Department of Health and Human Services. PAR-16-202: improving smoking cessation in socioeconomically disadvantaged populations via scalable interventions (R01). Bethesda, MD: National Cancer Institute, National Institutes of Health; 2016. URL: https://grants.nih.gov/grants/guide/pa-files/PAR-16-202.html [accessed 2018-07-23] [WebCite Cache ID 718BBMNTI]
10. American Lung Association. Cutting tobacco's rural roots: tobacco use in rural communities. Washington, DC: American Lung Association; 2015. URL: http://www.lung.org/our-initiatives/research/lung-health-disparities/tobacco-in-rural-communities.html [accessed 2018-07-23] [WebCite Cache ID 718BFFqp5]
11. Ziller E. Rural Populations, Culture, and Tobacco Use. South Portland, ME: Maine Center for Disease Control and Prevention; 2017.
12. National Center for Health Statistics. Health, United States, 2016: with chartbook on long-term trends in health. Hyattsville, MD: National Center for Health Statistics; 2017. URL: https://www.cdc.gov/nchs/data/hus/hus16.pdf [accessed 2018-07-23] [WebCite Cache ID 718BKlBYb]
13. Graham AL, Amato MS. Twelve million smokers look online for smoking cessation help annually: Health Information National Trends Survey data, 2005–2017. Nicotine Tob Res 2018 Apr 11:100-111. [doi: 10.1093/ntr/nty043] [Medline: 29660037]
14. Graham AL, Carpenter KM, Cha S, Cole S, Jacobs MA, Raskob M, et al. Systematic review and meta-analysis of Internet interventions for smoking cessation among adults. Subst Abuse Rehabil 2016;7:55-69 [FREE Full text] [doi: 10.2147/SAR.S101660] [Medline: 27274333]
15. Taylor GMJ, Dalili MN, Semwal M, Civljak M, Sheikh A, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2017 Sep 04;9:CD007078. [doi: 10.1002/14651858.CD007078.pub5] [Medline: 28869775]
16. BecomeAnEX.org. Stop Smoking Start Living!. Washington, DC: Truth Initiative; 2018. URL: https://excommunity.becomeanex.org/ [accessed 2018-10-01] [WebCite Cache ID 72qhbwXH2]
17. Graham AL, Cobb CO, Cobb NK. The internet, social media, and health decision-making. In: Diefenbach MA, Miller-Halegoua S, Bowen DJ, editors. Handbook of Health Decision Science. New York, NY: Springer; 2016:335-355.
18. Internet/broadband fact sheet. Washington, DC: Pew Research Center; 2018 Feb 05. URL: http://www.pewinternet.org/fact-sheet/internet-broadband/ [accessed 2018-07-23] [WebCite Cache ID 718BZsHTj]
19. 2016 broadband progress report. Washington, DC: Federal Communications Commission; 2016 Jan 29. URL: https://www.fcc.gov/reports-research/reports/broadband-progress-reports/2016-broadband-progress-report [accessed 2018-07-23] [WebCite Cache ID 718Bg43Sl]
20. Danaher BG, Hart LG, McKay HG, Severson HH. Measuring participant rurality in Web-based interventions. BMC Public Health 2007 Aug 31;7:228 [FREE Full text] [doi: 10.1186/1471-2458-7-228] [Medline: 17764564]
21. Campbell HS, Baskerville NB, Hayward LM, Brown KS, Ossip DJ. The reach ratio--a new indicator for comparing quitline reach into smoking subgroups. Nicotine Tob Res 2014 Apr;16(4):491-495 [FREE Full text] [doi: 10.1093/ntr/ntt192] [Medline: 24311698]
22. Baskerville NB, Brown KS, Nguyen NC, Hayward L, Kennedy RD, Hammond D, et al. Impact of Canadian tobacco packaging policy on use of a toll-free quit-smoking line: an interrupted time-series analysis. CMAJ Open 2016;4(1):E59-E65 [FREE Full text] [doi: 10.9778/cmajo.20150104] [Medline: 27280115]
23. Amato MS, Boyle RG, Lien B. Use of tobacco and tobacco cessation services in urban and rural Minnesota. 2016 Presented at: Society for Research and Nicotine 22nd Annual Meeting; Mar 2-5, 2016; Chicago, IL, USA.
24. McCausland KL, Curry LE, Mushro A, Carothers S, Xiao H, Vallone DM. Promoting a web-based smoking cessation intervention: implications for practice. Cases Public Health Commun Mark 2011;5:3-26.
25. Fiore MC, Jaén CR, Baker TB, Bailey WC, Benowitz NL, Curry SJ, Tobacco Use and Dependence Guidelines Panel. Treating tobacco's use and dependence: 2008 update. Clinical practice guideline. Rockville, MD: US Department of Health and Human Services, Public Health Service; 2008 May. URL: https://bphc.hrsa.gov/buckets/treatingtobacco.pdf [accessed 2018-10-01] [WebCite Cache ID 72qjC21ov]
26. Economic Research Service. Documentation: 2013 Rural-Urban Continuum Codes. Washington, DC: Economic Research Service, United States Department of Agriculture; 2017 Nov 27. URL: https://www.ers.usda.gov/data-products/rural-urban-continuum-codes/documentation/ [accessed 2018-07-23] [WebCite Cache ID 718C4oxuk]
27. Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality. 2016 National Survey On Drug Use And Health methodological summary and definitions. Rockville, MD: SAMHSA; 2017 Sep.

URL: https://www.samhsa.gov/data/sites/default/files/NSDUH-MethodSummDefs-2016/NSDUH-MethodSummDefs-2016.pdf [accessed 2018-07-23] [WebCite Cache ID 718CCClGF]

28. Zients J. Revised delineations of metropolitan statistical areas, micropolitan statistical areas, and combined statistical areas, and guidance on uses of the delineations of these areas. Washington, DC: Office of Management and Budget, Executive Office of the President; 2013 Feb 28. URL: https://obamawhitehouse.archives.gov/sites/default/files/omb/bulletins/2013/b-13-01.pdf [accessed 2018-08-20] [WebCite Cache ID 71ouf5eSE]
29. Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality. Results from the 2013 National Survey on Drug Use and Health: detailed tables. Rockville, MD: SAMHSA; 2014 Sep 04. URL: https://www.samhsa.gov/data/sites/default/files/NSDUH-DetTabs2013.pdf [accessed 2018-07-23] [WebCite Cache ID 718CLD0X6]
30. Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality. Results from the 2014 National Survey on Drug Use and Health: detailed tables. Rockville, MD: SAMHSA; 2015 Sep 10. URL: https://www.samhsa.gov/data/sites/default/files/NSDUH-DetTabs2014/NSDUH-DetTabs2014.pdf [accessed 2018-07-23] [WebCite Cache ID 718CQYqx9]
31. Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality. Results from the 2016 National Survey on Drug Use and Health: detailed tables. Rockville, MD: SAMHSA; 2017 Sep 07. URL: https://www.samhsa.gov/data/sites/default/files/NSDUH-DetTabs-2016/NSDUH-DetTabs-2016.pdf [accessed 2018-07-23] [WebCite Cache ID 718CTzRcG]
32. Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality. 2017 National Survey on Drug Use and Health: detailed tables. Rockville, MD: SAMHSA; 2018. URL: https://www.samhsa.gov/data/report/2017-nsduh-detailed-tables [accessed 2018-10-09] [WebCite Cache ID 732gh7QMS]
33. Fagerland MW, Lydersen S, Laake P. Recommended confidence intervals for two independent binomial proportions. Stat Methods Med Res 2015 Apr;24(2):224-254. [doi: 10.1177/0962280211415469] [Medline: 21996567]
34. Kingsbury JH, Parks MJ, Amato MS, Boyle RG. Deniers and admitters: examining smoker identities in a changing tobacco landscape. Nicotine Tob Res 2016 Dec;18(11):2130-2137. [doi: 10.1093/ntr/ntw110] [Medline: 27085084]
35. Schane RE, Glantz SA, Ling PM. Nondaily and social smoking: an increasingly prevalent pattern. Arch Intern Med 2009 Oct 26;169(19):1742-1744 [FREE Full text] [doi: 10.1001/archinternmed.2009.315] [Medline: 19858429]
36. Hart LG, Larson EH, Lishner DM. Rural definitions for health policy and research. Am J Public Health 2005 Jul;95(7):1149-1155. [doi: 10.2105/AJPH.2004.042432] [Medline: 15983270]

Abbreviations

NSDUH: National Survey of Drug Use and Health
ReRa: reach ratio
RUCC: Rural-Urban Continuum Codes
ZCTA: ZIP Code Tabulation Area

Edited by G Eysenbach; submitted 23.07.18; peer-reviewed by N Baskerville, K Bold; comments to author 16.08.18; revised version received 23.08.18; accepted 10.09.18; published 24.10.18.

Please cite as:
Amato MS, Graham AL
Geographic Representativeness of a Web-Based Smoking Cessation Intervention: Reach Equity Analysis
J Med Internet Res 2018;20(10):e11668
URL: http://www.jmir.org/2018/10/e11668/
doi:10.2196/11668
PMID:30355557

Review

Caregiver-Focused, Web-Based Interventions: Systematic Review and Meta-Analysis (Part 2)

Jenny Ploeg[1,2,3,4*], RN, PhD; Muhammad Usman Ali[5*], MSc, MD; Maureen Markle-Reid[1,3,4*], RN, MScN, PhD; Ruta Valaitis[1,3,4,6,7*], RN, PhD; Amy Bartholomew[1*], RN, MScN; Donna Fitzpatrick-Lewis[5*], MSW; Carrie McAiney[8*], PhD; Diana Sherifali[4,9,10], PhD, RN, BScN, CDE

[1]Aging, Community and Health Research Unit, School of Nursing, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada
[2]Department of Health, Aging and Society, McMaster University, Hamilton, ON, Canada
[3]McMaster Institute for Research on Aging, McMaster University, Hamilton, ON, Canada
[4]Faculty of Health Sciences, School of Nursing, McMaster University, Hamilton, ON, Canada
[5]Department of Health Research Methods, Evidence and Impact, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada
[6]World Health Organization Collaborating Centre for Primary Care and Health Human Resources, Hamilton, ON, Canada
[7]Department of Family Medicine, McMaster University, Hamilton, ON, Canada
[8]Program for Interpersonal Practice, Education and Research, Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, ON, Canada
[9]McMaster Evidence Review and Synthesis Team, School of Nursing, Faculty of Health Sciences, McMaster University, Hamilton, ON, Canada
[10]Diabetes Care and Research Program, Hamilton Health Sciences, Hamilton, ON, Canada
[*]these authors contributed equally

Corresponding Author:
Diana Sherifali, PhD, RN, BScN, CDE
Faculty of Health Sciences
School of Nursing
McMaster University
1280 Main Street West, HSC 3N28F
Hamilton, ON, L8S 4K1
Canada
Phone: 1 905 525 9140 ext 21435
Fax: 1 905 521 8834
Email: dsherif@mcmaster.ca

Abstract

Background: Approaches to support the health and well-being of family caregivers of adults with chronic conditions are increasingly important given the key roles caregivers play in helping family members to live in the community. Web-based interventions to support caregivers have the potential to lessen the negative health impacts associated with caregiving and result in improved health outcomes.

Objective: The primary objective of this systematic review and meta-analysis was to examine the effect of caregiver-focused, Web-based interventions, compared with no or minimal Web-based interventions, on caregiver outcomes. The secondary objective was to assess the effect of different types of Web-based interventions (eg, education, peer and professional psychosocial support, and electronic monitoring of the care recipient), compared with no or minimal Web-based interventions, on caregiver outcomes.

Methods: MEDLINE, EMBASE, CIHAHL, PsychInfo, Cochrane, and AgeLine were searched from January 1995 to April 2017 for relevant randomized controlled trials (RCTs) or controlled clinical trials (CCTs) that compared caregiver-focused, Web-based intervention programs with no or minimal Web-based interventions for caregivers of adults with at least one chronic condition. Studies were included if they involved: adult family or friend caregivers (aged ≥18 years) of adults living in the community with a chronic condition; a caregiver-focused, Web-based intervention of education or psychosocial support or electronic monitoring of the care recipient; and general caregiver outcomes (ie, burden, life satisfaction, self-efficacy or mastery, reaction to problem behavior, self-esteem, strain, and social support). Title and abstract as well as full-text screening were completed in duplicate. Data were extracted by a single reviewer and verified by a second reviewer, and risk of bias assessments were completed accordingly. Where possible, data for these caregiver outcomes were meta-analyzed.

Results: The search yielded 7927 unique citations, of which 294 studies were screened at full text. Of those, 14 studies met the inclusion criteria; 12 were RCTs and 1 study was a CCT. One study used an RCT design in 1 country and a CCT design in 2 other countries. The beneficial effects of any Web-based intervention program, compared with no or minimal Web-based intervention, resulted in a mean increase of 0.85 points (95% CI 0.12 to 1.57) for caregiver self-esteem, a mean increase of 0.36 points (95% CI 0.11 to 0.62) for caregiver self-efficacy or mastery, and a mean decrease of 0.32 points (95% CI −0.54 to −0.09) for caregiver strain. However, the results are based on poor-quality studies.

Conclusions: The review found evidence for the positive effects of Web-based intervention programs on self-efficacy, self-esteem, and strain of caregivers of adults living with a chronic condition. Further high-quality research is needed to inform the effectiveness of specific types of Web-based interventions on caregiver outcomes.

Trial Registration: PROSPERO CRD42018091715; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=91715 (Archived by WebCite at http://www.webcitation.org/738zAa5F5)

(J Med Internet Res 2018;20(10):e11247) doi:10.2196/11247

KEYWORDS

burden; caregivers; chronic conditions; education; internet; meta-analysis; support; Web-based interventions

Introduction

The number of individuals living with chronic conditions is on the rise globally [1]. Family and friend caregivers provide up to 75% of the health and supportive care needs for older adults living in the community in Canada [2]. Although caregiving can be very rewarding, it is also associated with adverse physical, mental, and psychosocial health outcomes [3-5]. Examples of negative outcomes as a result of caregiving include burden, strain, being dissatisfied with life, feeling alone or isolated, and having low self-efficacy [3-8]. Thus, practical solutions to address the needs of caregivers are urgently needed.

Recently, there has been great interest in the use of Web-based interventions to support caregivers. It has been suggested that the delivery of health care interventions through the Web may result in improved accessibility of services as well as reduced health care costs [9]. There is accumulating evidence for the positive effect of caregiver-focused, Web-based interventions in 11 recent systematic or narrative reviews [10-20]. All of these reviews provided some evidence of improvements in caregivers' health or well-being (eg, burden, depression, self-efficacy, and confidence) as a result of Web-based programs. Most reviews included studies with both high- and low-quality designs and noted the limited methodological quality of included studies as a concern. In addition, most reviews did not examine the effect of different types of Web-based support on caregivers. Finally, none of the reviews included a meta-analysis to quantify the magnitude of effect across studies.

The primary objective of this study was to conduct a systematic review and meta-analysis to assess the effect of caregiver-focused, Web-based interventions, compared with no or minimal web-based interventions, on outcomes for caregivers of adults with at least one chronic condition living in the community. The caregiver outcomes examined in this paper include burden, life satisfaction, self-efficacy or mastery, reaction to problem behavior, self-esteem, strain, and social support. The secondary objective was to examine whether specific types of Web-based interventions had a beneficial effect on these caregiver outcomes, to address previous review limitations. Of note, this review included only studies with the most rigorous designs, randomized controlled trials (RCTs) and controlled clinical trials (CCTs). This is a companion paper to a systematic review and meta-analysis that examines the effect of internet-based interventions on caregiver mental health outcomes [21].

Methods

Reporting Guidelines

This systematic review and meta-analysis was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analysis guidelines [22].

Population

The population of interest included family and friend caregivers, aged ≥18 years, who were providing caregiving support to adults (≥18 years) living in the community with at least one chronic condition (ie, "care recipient").

Interventions

Studies selected for this systematic review included those that examined any caregiver-focused, Web-based modality to deliver an intervention, which could include either a single component program or multimodal program.

Outcomes

The outcomes assessed in this meta-analysis included the following caregiver outcomes: burden, life satisfaction, self-efficacy or mastery, reaction to problem behavior, self-esteem, strain, and social support. Mental health outcomes are addressed in a companion paper [21].

Study Design

Inclusion and Exclusion

Studies were included if they met the following inclusion criteria: study designs were an RCT or CCT; studies examined any Web-based intervention program for caregivers of older adults having at least one chronic condition and living in the community; studies were published between January 1, 1995 and April 19, 2017; studies were published in English; studies reported on at least one caregiving outcome of interest (burden, life satisfaction, self-efficacy or mastery, reaction to problem behavior, self-esteem, strain, or social support); studies used any measurement tool to examine the outcomes of interest; and studies in which the control group received no or minimal Web-based intervention. Of note, there were no restrictions on the nature of chronic conditions of care recipients. The exclusion criteria included all other types of study designs (ie, observational studies and case reports), gray or unpublished literature, conference abstracts, and letters or editorials. Furthermore, all published study protocols without preliminary results for data extraction were also excluded.

Search Strategy

A peer-reviewed search strategy was developed by 2 research librarians at McMaster University. EMBASE, MEDLINE, PsychInfo, CINAHL, Cochrane, and AgeLine were searched for studies published between January 1, 1995 and April 19, 2017. In addition, reference lists of systematic reviews were searched for relevant studies not captured by the initial search. Results were deduplicated, and the citations were uploaded to a secure Web-based platform. Multimedia Appendix 1 provides detailed information about the search terms.

Selection of Studies

Two reviewers independently selected studies for possible inclusion based on a title and abstract review. Studies meeting the inclusion criteria by either reviewer then underwent full-text review. Any disagreements were discussed between reviewers, and a third party was involved to help reach consensus, as necessary.

Data Extraction and Quality Assessment

Full-data extraction, including characteristics of included studies, was completed by one reviewer and verified by a second reviewer. The risk of bias (RoB) found in individual studies was assessed by one reviewer and verified by a second reviewer. RoB was assessed using the Cochrane RoB framework [23], which evaluates the level of bias for sequence generation, allocation concealment, blinding, completeness of outcome assessment, selective reporting, and other biases. The quality of the clinical evidence was critically appraised by one reviewer and verified by a second reviewer using the Grading of Recommendations Assessment, Development, and Evaluation system (GRADE), which evaluates the risk for bias, inconsistency, indirectness, and imprecision for each outcome [24]. Disagreements were resolved through consensus between the 2 reviewers.

Data Analysis

We used a meta-analysis to combine the results across studies for each outcome using the published data from included studies. To perform the meta-analysis, we used immediate posttreatment data (mean, SD) for continuous outcomes such as burden, life satisfaction, self-efficacy or mastery, reaction to problem behavior, self-esteem, strain, and social support. In addition, we used intention-to-treat outcome data where possible; however, if no intention-to-treat data were reported, we used outcome data obtained from those who completed the study.

The DerSimonian and Laird random-effects models with the inverse variance method were used to generate the summary measures of effect in the form of standardized mean difference (SMD) [25]. SMD accounts for similar outcomes measured using different assessment tools (eg, caregiver burden assessed using different outcome measures such as the Zarit Burden Interview and Caregiver Quality of Life Scale). In this situation, it was necessary to standardize the results of the studies to a uniform scale before they could be combined in quantitative synthesis. SMDs were calculated using change from the baseline data for intervention and control groups for each study with relevant outcome data. For each outcome, data from the corresponding study were used to calculate the mean difference between pretreatment (baseline) and posttreatment (final or endpoint) values along with its SD for both intervention and control groups. In studies where SD was not reported, we calculated it from the reported SE of the mean, 95% CIs and *P* values, or *z* values using equations provided in Chapters 7 and 9 of the Cochrane Handbook for Systematic Reviews of Interventions [26,27]. SMD is interpreted on the basis of its magnitude according to Cohen *d* recommended thresholds (~0.2, small effect; ~0.5, medium effect; and ~0.8, large effect) [28].

The primary meta-analysis examined caregiver-focused, Web-based interventions by caregiver outcome. Subsequently, the secondary meta-analysis examined the effects of specific types of caregiver-focused, Web-based intervention programs on caregiver outcomes. Based on our previous work [18], intervention types were categorized accordingly: Web-based information or education only; Web-based information or education plus peer psychosocial support; Web-based information or education plus professional psychosocial support; Web-based information or education plus combined peer and professional psychosocial support; and Web-based information or education plus professional psychosocial support plus electronic monitoring of the care recipient.

The statistical heterogeneity of combined studies was examined using standard methods. The I^2 statistic was used to quantify the magnitude of statistical heterogeneity between studies where I^2 of 30%-60% represents moderate and I^2 of >60% represents substantial heterogeneity [26]. We used $P<.10$ as a guide to indicate where statistically significant heterogeneity may exist, upon which a closer examination of study differences was performed. All analyses were performed using Review Manager (RevMan Version 5.3, The Nordic Cochrane Centre, The Cochrane Collaboration, Copenhagen, Denmark) [29], STATA (version 14; Stata Corp, College Station, Texas, USA) [30], and GRADEpro Guideline Development Tool software packages [31].

Results

Search

The search resulted in 7927 unique citations, which were screened independently by 2 project staff, as seen in Figure 1. At title and abstract screening, we excluded 7633 studies, leaving 294 studies to be screened at full text. Of these, we identified 14 studies (16 papers) that met the inclusion criteria for this review. References lists of the on-topic systematic reviews and included studies were searched, but no additional studies were added.

Summary of Included Studies

Multimedia Appendix 2 presents the purpose, methods, participants, and intervention of the included studies.

Study Design

Type of Studies

Among 14 included studies, 12 were RCTs [32-43], 1 was a CCT [44], and 1 study used an RCT design in 1 country and a CCT design in 2 other countries [45]. Companion papers [46,47]were included for the studies by DuBenske [37] and Smith [41] respectively. Of 12 RCTs, 4 were conducted in Europe [33-36], 7 in the United States [32,37,38,40-43], and 1 in South Korea [39]. The one CCT was conducted across the United States, Puerto Rico, and Mexico [44], and the study that used both CCT and RCT designs was conducted across 3 European countries [45]. All included studies had relatively small sample sizes (≤150 subjects per arm) and most had a length of follow-up of ≤6 months. One study included a slightly longer study follow-up period of 1 year [42]. In addition, 7 of 14 studies included reference to a theoretical or conceptual framework for the intervention, including stress and coping [34,35,37,41,43], framework of systemic organization [42], and the concept of ambient assisted living [33].

Study Population

Most studies included caregivers aged ≥50 years (mean age ranged from 53.8-66.0) [32,34-39,41,42,45], except 1 study that included caregivers who were working and reported a slightly lower mean age of 46.9 years [43]. In addition, 2 studies did not provide information on the average age of caregivers [33,44], and 1 study reported that 40% were >50 years [40]. Next, 11 of 14 studies reported caregiver gender; in 10 studies, more than half of the caregivers were females (56.3%-100%). In relation to the type of chronic conditions among care recipients, 9 studies included persons with some form of dementia [32-36,38,43-45]. In 3 studies, care recipients were stroke survivors [39,41,42]. Care recipients in 1 study had nonsmall cell lung cancer [37] and in another study, care recipients had brain injury [40].

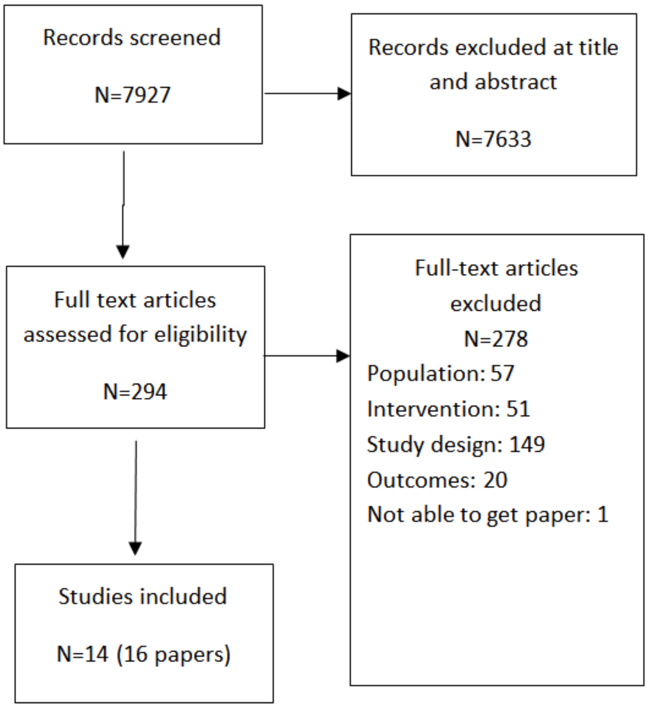
Type of Web-Based Intervention

Among 14 included studies, 3 studies used a *Web-based information or education only* intervention [38,40,43], 3 studies used a *Web-based information or education plus peer psychosocial support* intervention [33-35], 1 study used a *Web-based information or education plus professional psychosocial support* intervention [39], 6 studies used a *Web-based information or education plus combined peer and professional psychosocial support* intervention [32,36,37,41,42,44], and 1 study used a *Web-based information or education plus professional psychosocial support intervention plus electronic monitoring* [45].

Comparator Groups

The comparator groups received usual care or were part of a wait-list control wherein they had access to the Web-based program at the end of the study [32-36,39,42,43,45], had access to printed materials [44], or had access to a website with general information related to the condition or resources [37,38,40,41].

Figure 1. Flow diagram for the study selection of Web-based interventions on caregiving outcomes.



Outcomes

Among 14 included studies, outcomes examined included the following: burden (n=5); life satisfaction (n=3); self-efficacy or mastery (n=9); reaction to problem behavior (n=2); self-esteem (n=1); strain (n=1); and social support (n=2). Measurement tools to assess caregiver outcomes varied across included studies (Multimedia Appendix 3).

Table 1 shows the results of the critical appraisal of individual studies for the level of bias for sequence generation, allocation concealment, blinding, completeness of outcome assessment, selective reporting, and other biases. Overall, the Cochrane RoB showed the mixed quality of study methodology; 1 study had low RoB [35], 4 studies had high RoB [34,37,44,45], and 9 studies had unclear RoB because of the lack of relevant details in the published papers [32,33,36,38-43].

Effectiveness of Web-Based Interventions

The meta-analysis included an examination of any Web-based intervention as well as an examination of each type of Web-based intervention by caregiver outcome. Multimedia Appendix 4 shows all forest plots.

Any Web-Based Intervention

Table 2 summarizes the results of the meta-analysis of any caregiver-focused, Web-based intervention on caregiver outcomes. Compared with no or minimal Web-based intervention, any type of Web-based intervention resulted in a statistically significant mean increase of 0.85 points (95% CI 0.12 to 1.57) for caregiver self-esteem, 0.36 points (95% CI 0.11 to 0.62) for caregiver self-efficacy or mastery, and a decrease of 0.32 points (95% CI −0.54 to −0.09) for caregiver strain. There were no statistically significant differences between groups for the caregiver outcomes of caregiver burden, life satisfaction, reaction to problem behavior, and social support. In addition, heterogeneity for the combined effect estimate was observed for the outcomes of caregiver burden, self-efficacy or mastery, reaction to problem behavior, and social support. The overall GRADE quality of evidence for each outcome ranged from moderate to very low. See Multimedia Appendix 5 for the full GRADE assessment details.

Effect of Different Types of Web-Based Interventions

Caregiver outcomes of interest were examined for each type of Web-based intervention, as shown in Table 3. For *information or education only* interventions, results showed a significant reduction with small effect sizes in caregiver strain (1 study; SMD=−0.32, 95% CI −0.54 to −0.09, *P*=.007) and self-efficacy or mastery (1 study; SMD=0.31, 95% CI 0.08 to 0.53, *P*=.009). These results were based on the moderate quality of evidence. The remaining outcomes of life satisfaction and reaction to problem behavior, which were assessed in only one study each, did not show statistically significant differences between groups.

For studies that examined *information or education plus peer psychosocial support*, there were no differences between intervention and control groups for any of the outcomes, including burden, life satisfaction, self-efficacy or mastery, and reaction to problem behavior; the quality of this evidence was very low. For studies that examined *information or education plus professional psychosocial support*, results showed a mean increase of 1.2 points (95% CI 0.48 to 1.92) for self-efficacy or mastery compared with no or minimal Web-based intervention; the quality of this evidence was very low.

Table 1. Risk of bias (RoB) of included studies.

Author, year	Sequence Generation	Allocation Concealment	Blinding of Participants/ Providers	Blinding of Outcome Assessment	Incomplete Outcome Data	Selective Reporting	Other Bias	Overall RoB
Beauchamp et al, 2005 [43]	Unclear	Unclear	Unclear	Unclear	Low	Low	Low	Unclear
Cristancho-Lacroix et al, 2015 [34]	Low	Unclear	High	High	Low	Low	High	High
DuBenske et al, 2014 [37] (Companion Paper: Gustafson et al, 2013 [46])	Unclear	Unclear	High	High	High	Low	High	High
Hattink et al, 2015 [35]	Low	Unclear	Low	Low	Low	Low	Low	Low
Hattink et al, [45]	High	High	Unclear	Unclear	Low	Low	Low	High
Kajiyama et al, 2013 [38]	Unclear	Unclear	Unclear	Unclear	High	Low	Low	Unclear
Kim et al, 2013 [39]	Low	Unclear	High	Unclear	Low	Low	Low	Unclear
McLaughlin et al, 2013 [40]	Unclear	Unclear	Unclear	Unclear	Low	Low	Low	Unclear
Fowler et al, 2016 [32]	Low	Unclear	High	Unclear	Low	Low	Low	Unclear
Núñez-Naveira et al, 2016 [33]	Low	Unclear	Unclear	Unclear	Low	Low	Low	Unclear
Pagán-Ortiz et al, 2014 [44]	High	High	Unclear	Unclear	Low	Low	Unclear	High
Pierce et al, 2009 [42]	Unclear	Unclear	Unclear	Unclear	High	Low	Low	Unclear
Smith et al, 2012 [41] (Companion Paper: Steiner et al, 2002 [47])	Low	Unclear	Unclear	Low	Low	Low	Low	Unclear
Torkamani et al, 2014 [36]	Unclear	Unclear	Unclear	Unclear	Low	Low	Low	Unclear

Table 2. A summary of the effectiveness of any Web-based intervention.

Caregiver outcomes	Number of studies	Intervention, n	Control, n	Estimate standardized mean difference (95% CI)	I^2 (%)	Grading[a]
Caregiver burden	5	132	147	0.03 (−0.31 to 0.36)	48	Very low
Life satisfaction	3	170	165	−0.17 (−0.39 to 0.04)	0	Very low
Self-efficacy or mastery	9	306	309	0.36 (0.11 to 0.62)	46	Low
Reaction to problem behavior	2	71	81	−0.10 (−0.66 to 0.45)	63	Very low
Self-esteem	1	15	17	0.85 (0.12 to 1.57)	N/A[b]	Very low
Caregiver strain	1	150	149	−0.32(−0.54 to −0.09)	N/A	Moderate
Social support	2	30	34	−0.38 (−1.12 to 0.35)	53	Very low

[a] Grading of recommendations assessment, development, and evaluation system quality assessment.

[b] N/A: not applicable.

Table 3. A summary of the effectiveness of types of Web-based interventions.

Caregiver outcomes	Number of studies	Intervention, n	Control, n	Estimate standard mean difference (95% CI)	I^2 (%)	Grading[a]
Information or education						
Life satisfaction	1	104	97	−0.22 (−0.50 to 0.06)	N/A[b]	Very low
Self-efficacy or mastery	1	150	149	0.31 (0.08 to 0.53)	N/A	Moderate
Reaction to problem behavior	1	46	57	−0.35 (−0.75 to 0.04)	N/A	Very low
Strain	1	150	149	−0.32 (−0.54 to −0.09)	N/A	Moderate
Information or education plus peer psychosocial support						
Burden	2	46	49	0.17 (−0.24 to 0.57)	0	Very low
Life satisfaction	1	30	31	0.08 (−0.43 to 0.58)	N/A	Very low
Self-efficacy or mastery	3	76	80	0.14 (−0.41 to 0.69)	66	Very low
Reaction to problem behavior	1	25	24	0.22 (−0.34 to 0.78)	N/A	Very low
Information or education plus professional psychosocial support						
Self-efficacy or mastery	1	18	18	1.20 (0.48-1.92)	N/A	Very low
Information or education plus peer and professional psychosocial support						
Burden	3	86	98	−0.03 (−0.57 to 0.50)	67	Very low
Life satisfaction	1	36	37	−0.24 (−0.70 to 0.22)	N/A	Very low
Self-efficacy or mastery	3	45	47	0.52 (0.10-0.94)	0	Very low
Self-esteem	1	15	17	0.85 (0.12-1.57)	N/A	Very low
Social support	2	30	34	−0.38 (−1.12 to 0.35)	53	Very low
Information or education plus professional psychosocial support plus monitoring						
Self-efficacy or mastery	1	17	15	0.17 (−0.52 to 0.87)	N/A	Very low

[a] Grading of recommendations assessment, development, and evaluation system quality assessment.

[b] N/A: not applicable.

For studies that examined *information or education plus combined peer and professional psychosocial support*, results showed a mean increase of 0.85 points (95% CI 0.12 to 1.57) for self-esteem and 0.52 points (95% CI 0.10 to 0.94) for self-efficacy or mastery compared with no or minimal Web-based intervention; the quality of this evidence was very low. For the outcomes of burden, life satisfaction, and social support, there were no statistically significant differences between groups. Finally, the single study that examined *information or education plus professional psychosocial support plus electronic monitoring* found no statistically significant difference between groups for the outcome of self-efficacy or mastery; the quality of this evidence was very low.

Discussion

Principal Findings

To the best our knowledge, this paper and its companion paper, focused on caregiver mental health outcomes [21], are the first meta-analyses examining the effect of caregiver-focused, Web-based interventions on outcomes of caregivers of adults with chronic conditions living in the community. This systematic review and meta-analysis showed small to medium beneficial effects of Web-based interventions on caregiver outcomes of self-esteem, self-efficacy, or mastery and strain but no effect on the burden, life satisfaction, reaction to problem behavior, and social support. For Web-based information or education interventions, there was a small effect size on self-efficacy or mastery and strain with a moderate quality of evidence. For Web-based information or education plus professional psychosocial support (1 study), there was a large effect size for self-efficacy or mastery, but the quality of evidence was very low. For Web-based information or education plus combined peer and professional psychosocial support, there was a large effect size for self-esteem and moderate effect size for self-efficacy or mastery, but the quality of the evidence was also very low. Finally, for Web-based information or education plus professional psychosocial support plus electronic monitoring, there was no effect on self-efficacy or mastery.

There are a number of possible reasons why consistent findings across caregiver outcomes were not shown. According to the GRADE scores, the quality of evidence was low or very low for most of the outcomes examined, and none of the outcomes was rated as having high-quality evidence. Furthermore, some outcomes were assessed in only a single study; there was variability in the assessment tools used to assess outcomes, caregiver characteristics varied across studies, and very few studies examined different types of Web-based interventions, reflecting that this is an emerging area of research.

In relation to RoB, 4 studies had high RoB in the area of blinding participants or providers, 3 studies had high RoB in the area of incomplete outcome data, and 2 studies had high RoB for allocation concealment, blinding of outcome assessment, and sequence generation. In addition, there were many areas where RoB could not be determined because of the lack of information in the published papers; for example, RoB was unclear in 12 of 14 studies related to the allocation concealment, in 10 of 14 studies related to blinding of the outcome assessment, and in 9 of 14 studies related to blinding of participants and providers. It is vital that the authors of such trials provide more detailed information about trial procedures using the Consolidated Standards of Reporting Trials guidelines for nonpharmacological interventions [48] because this would enable more accurate assessment of studies for bias and may help to improve the quality of evidence in this area.

The improvements in caregiver self-efficacy or mastery as a result of Web-based interventions in this study are promising. These improvements were observed in Web-based interventions that included information or education in combination with either professional psychosocial support or both peer and professional psychosocial support. Caregiver self-efficacy, or a person's perception of their ability to perform tasks related to caregiving competently, is a modifiable factor that is important in understanding the effect of Web-based caregiver interventions. Previous research has shown that higher self-efficacy is associated with fewer depressive symptoms among dementia family caregivers [49,50]. Moreover, research suggests that self-efficacy for managing dementia may protect caregivers against burden and depression [51] and that it plays a mediating role between social support and depressive symptoms [52].

For the outcome of self-efficacy or mastery, the addition of professional psychosocial support to information or education only resulted in an increase in SMD from 0.31 to 1.20, whereas the addition of both peer and professional psychosocial support resulted in an increase in SMD from 0.31 to 0.52; this suggests that human support (either professional or peer or a combination of these), as previously shown in the review by Guay et al [13], plays an important role in improving caregiver outcomes.

The companion paper [21], a meta-analysis of the effect of internet-based interventions on caregiver mental health, showed that such interventions also result in a reduction in depressive symptoms, stress or distress, and anxiety. Because the companion paper included many of the same studies in this meta-analysis, it had similar limitations in relation to the quality of the evidence.

A theoretical basis for Web-based interventions has been shown to be effective [13,53]. Half of the included studies reported using theories, such as cognitive theories of stress, to develop their interventions [34,41,43]. Interventions that included behavioral change techniques, such as stress management, may have contributed to significant findings. It is recommended that Web-based interventions for caregivers include a strong theoretical base [54] and include strategies to support improved self-efficacy, stress management, and coping.

Strengths and Limitations

This review uses meta-analysis to summarize the most relevant trial evidence available on the effects of Web-based interventions on caregiving outcomes. Another strength of this review is the *a priori* selection of rigorous methodological designs, including only RCTs and CCTs. The review was conducted using a comprehensive search strategy and methodologically rigorous processes for systematic reviews and meta-analyses. The included papers were published between 2005 and 2016 with 12 of 14 published since 2012, which reflects the growing interest in Web-based technology to support caregivers. One of the limitations of the review involved the overall low quality of the studies included, despite being RCTs and CCTs. Owing to the considerable heterogeneity of interventions across studies, results were examined according to the types of Web-based interventions. However, there were very few studies that used each type of Web-based intervention across the outcomes of interest.

Conclusions

This paper and its companion paper [21] are the first meta-analyses of the effect of Web-based interventions for caregivers of community-living adults with chronic conditions on caregiver outcomes. The findings indicate that there is accumulating evidence for the positive effect of caregiver-focused, Web-based interventions to support family and friend caregivers. However, future high-quality research with stronger study designs, larger sample sizes, and the use of standardized tools to facilitate meta-analysis and assessment of clinical relevance are needed to understand the effect of such interventions, particularly multicomponent interventions using peer or professional support.

Acknowledgments

This work is part of a program of research (Aging, Community and Health Research Unit) supported by the Canadian Institutes of Health Research Signature Initiative in Community-Based Primary Healthcare [55] (Funding Reference Number TTF 128261). This work was also supported by the McMaster Evidence Review and Synthesis team. MMR was supported through a Canada Research Chair in Person-Centred Interventions for Older Adults with Multimorbidity and their caregivers. RV was supported through the Dorothy C Hall Chair in Primary Health Care Nursing, School of Nursing, McMaster University.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms.
[PDF File (Adobe PDF File), 50 KB - jmir_v20i10e11247_app1.pdf]

Multimedia Appendix 2
Detailed characteristics of included studies.
[PDF File (Adobe PDF File), 111 KB - jmir_v20i10e11247_app2.pdf]

Multimedia Appendix 3
Caregiver outcomes and measurement assessment tools.
[PDF File (Adobe PDF File), 38 KB - jmir_v20i10e11247_app3.pdf]

Multimedia Appendix 4
Meta-analysis and forest plots.
[PDF File (Adobe PDF File), 120 KB - jmir_v20i10e11247_app4.pdf]

Multimedia Appendix 5
GRADE assessment details.
[PDF File (Adobe PDF File), 175 KB - jmir_v20i10e11247_app5.pdf]

References

1. World Health Organization. Non-communicable diseases. 2017. URL: http://www.who.int/en/news-room/fact-sheets/detail/noncommunicable-diseases
2. Health Council of Canada. 2012. Seniors in need, caregivers in distress: What are the home care priorities for seniors in Canada. www.carp URL: http://www.carp.ca/wp-content/uploads/2012/04/HCC_HomeCare_2d.pdf [WebCite Cache ID 72VVKW7BR]
3. Pinquart M, Sörensen S. Differences between caregivers and noncaregivers in psychological health and physical health: a meta-analysis. Psychol Aging 2003 Jun;18(2):250-267. [Medline: 12825775]
4. Roth DL, Fredman L, Haley WE. Informal caregiving and its impact on health: a reappraisal from population-based studies. Gerontologist 2015 Apr;55(2):309-319. [doi: 10.1093/geront/gnu177] [Medline: 26035608]
5. Turcotte M. Insights on Canadian society. Ottawa, ON; 2013. URL: https://www150.statcan.gc.ca/n1/pub/75-006-x/2013001/article/11858-eng.pdf [accessed 2018-09-17] [WebCite Cache ID 72VVYUzmf]
6. Adelman RD, Tmanova LL, Delgado D, Dion S, Lachs MS. Caregiver burden: a clinical review. JAMA 2014 Mar 12;311(10):1052-1060. [doi: 10.1001/jama.2014.304] [Medline: 24618967]
7. Anderson LA, Edwards VJ, Pearson WS, Talley RC, McGuire LC, Andresen EM. Adult caregivers in the United States: characteristics and differences in well-being, by caregiver age and caregiving status. Prev Chronic Dis 2013 Aug 15;10:E135 [FREE Full text] [doi: 10.5888/pcd10.130090] [Medline: 23948336]

8. Roth DL, Perkins M, Wadley VG, Temple EM, Haley WE. Family caregiving and emotional strain: associations with quality of life in a large national sample of middle-aged and older adults. Qual Life Res 2009 Aug;18(6):679-688 [FREE Full text] [doi: 10.1007/s11136-009-9482-2] [Medline: 19421895]
9. Eysenbach G. What is e-health? J Med Internet Res 2001 Jun;3(2):E20. [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
10. Boots LMM, de Vugt ME, van Knippenberg RJM, Kempen GIJM, Verhey FRJ. A systematic review of Internet-based supportive interventions for caregivers of patients with dementia. Int J Geriatr Psychiatry 2014 Apr;29(4):331-344. [doi: 10.1002/gps.4016] [Medline: 23963684]
11. Dam AEH, de Vugt ME, Klinkenberg IPM, Verhey FRJ, van Boxtel MP. A systematic review of social support interventions for caregivers of people with dementia: Are they doing what they promise? Maturitas 2016 Mar;85:117-130. [doi: 10.1016/j.maturitas.2015.12.008] [Medline: 26857890]
12. Egan KJ, Pinto-Bruno Á, Bighelli I, Berg-Weger M, van Straten A, Albanese E, et al. Online Training and Support Programs Designed to Improve Mental Health and Reduce Burden Among Caregivers of People With Dementia: A Systematic Review. J Am Med Dir Assoc 2018 Mar;19(3):200-206.e1. [doi: 10.1016/j.jamda.2017.10.023] [Medline: 29306605]
13. Guay C, Auger C, Demers L, Mortenson WB, Miller WC, Gélinas-Bronsard D, et al. Components and Outcomes of Internet-Based Interventions for Caregivers of Older Adults: Systematic Review. J Med Internet Res 2017 Sep 19;19(9):e313 [FREE Full text] [doi: 10.2196/jmir.7896] [Medline: 28928109]
14. Kaltenbaugh DJ, Klem ML, Hu L, Turi E, Haines AJ, Hagerty LJ. Using Web-based interventions to support caregivers of patients with cancer: a systematic review. Oncol Nurs Forum 2015 Mar;42(2):156-164. [doi: 10.1188/15.ONF.156-164] [Medline: 25806882]
15. Hu C, Kung S, Rummans TA, Clark MM, Lapid MI. Reducing caregiver stress with internet-based interventions: a systematic review of open-label and randomized controlled trials. J Am Med Inform Assoc 2015 Apr;22(e1):e194-e209. [doi: 10.1136/amiajnl-2014-002817] [Medline: 25125686]
16. Jackson D, Roberts G, Wu ML, Ford R, Doyle C. A systematic review of the effect of telephone, internet or combined support for carers of people living with Alzheimer's, vascular or mixed dementia in the community. Arch Gerontol Geriatr 2016;66:218-236. [doi: 10.1016/j.archger.2016.06.013] [Medline: 27372903]
17. Lee E. Do Technology-Based Support Groups Reduce Care Burden Among Dementia Caregivers? A Review. J Evid Inf Soc Work 2015;12(5):474-487. [doi: 10.1080/15433714.2014.930362] [Medline: 25794367]
18. Ploeg J, Markle-Reid M, Valaitis R, McAiney C, Duggleby W, Bartholomew A, et al. Web-Based Interventions to Improve Mental Health, General Caregiving Outcomes, and General Health for Informal Caregivers of Adults With Chronic Conditions Living in the Community: Rapid Evidence Review. J Med Internet Res 2017 Jul 28;19(7):e263 [FREE Full text] [doi: 10.2196/jmir.7564] [Medline: 28754652]
19. Tang WP, Chan CW, So WK, Leung DY. Web-based interventions for caregivers of cancer patients: A review of literatures. Asia Pac J Oncol Nurs 2014;1(1):9-15 [FREE Full text] [doi: 10.4103/2347-5625.135811] [Medline: 27981077]
20. Waller A, Dilworth S, Mansfield E, Sanson-Fisher R. Computer and telephone delivered interventions to support caregivers of people with dementia: a systematic review of research output and quality. BMC Geriatr 2017 Nov 16;17(1):265 [FREE Full text] [doi: 10.1186/s12877-017-0654-6] [Medline: 29145806]
21. Sherifali D, Ali M, Ploeg J, Markle-Reid M, Valaitis R, Bartholomew A, et al. Impact of Internet-Based Interventions on Caregiver Mental Health: Systematic Review and Meta-Analysis. J Med Internet Res 2018 Jul 03;20(7):e10668 [FREE Full text] [doi: 10.2196/10668] [Medline: 29970358]
22. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
23. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928 [FREE Full text] [Medline: 22008217]
24. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, GRADE Working Group. Grading quality of evidence and strength of recommendations. BMJ 2004 Jun 19;328(7454):1490 [FREE Full text] [doi: 10.1136/bmj.328.7454.1490] [Medline: 15205295]
25. DerSimonian R, Laird N. Meta-analysis in clinical trials. Controlled clinical trials 1986:177-188 [FREE Full text] [doi: 10.1016/0197-2456(86)90046-2]
26. Higgins J, Deeks JJ. Selecting studies and collecting data. In: Higgins J, Green S, editors. Cochrane handbook for systematic reviews of interventions: The Cochrane Collaboration; 2011.
27. Deeks JJ, Higgins J, Altman DG. Analysing data and undertaking meta-analyses. In: Deeks JJ, Higgins J, Altman DG, editors. Cochrane handbook for systematic reviews of interventions: The Cochrane Collaboration; 2011.
28. Cohen J. Statistical power analysis for the behavioral sciences. 2nd edition. Hillsdale, NJ: L. Erlbaum Associates; 1988:0805802835.
29. Review Manager (RevMan) [Computer Program]. Version 5.3. The Nordic Cochrane Centre, The Cochrane Collaboration: Copenhagen; 2011.
30. StataCorp. Stata statistical software. Station, TX: StataCorp LP; 2011:2-8.
31. GRADEpro [Computer program]. Hamilton, ON: McMaster University; 2015. URL: https://gradepro.org/

32. Fowler CN, Kott K, Wicks MN, Rutledge C. An Interprofessional Virtual Healthcare Neighborhood: Effect on Self-Efficacy and Sleep Among Caregivers of Older Adults With Dementia. J Gerontol Nurs 2016 Nov 01;42(11):39-47. [doi: 10.3928/00989134-20160901-02] [Medline: 27598270]
33. Núñez-Naveira L, Alonso-Búa B, de Labra C, Gregersen R, Maibom K, Mojs E, et al. UnderstAID, an ICT Platform to Help Informal Caregivers of People with Dementia: A Pilot Randomized Controlled Study. Biomed Res Int 2016;2016:5726465 [FREE Full text] [doi: 10.1155/2016/5726465] [Medline: 28116300]
34. Cristancho-Lacroix V, Wrobel J, Cantegreil-Kallen I, Dub T, Rouquette A, Rigaud A. A web-based psychoeducational program for informal caregivers of patients with Alzheimer's disease: a pilot randomized controlled trial. J Med Internet Res 2015;17(5):e117 [FREE Full text] [doi: 10.2196/jmir.3717] [Medline: 25967983]
35. Hattink B, Meiland F, van der Roest H, Kevern P, Abiuso F, Bengtsson J, et al. Web-Based STAR E-Learning Course Increases Empathy and Understanding in Dementia Caregivers: Results from a Randomized Controlled Trial in the Netherlands and the United Kingdom. J Med Internet Res 2015;17(10):e241 [FREE Full text] [doi: 10.2196/jmir.4025] [Medline: 26519106]
36. Torkamani M, McDonald L, Saez AI, Kanios C, Katsanou M, Madeley L, ALADDIN Collaborative Group. A randomized controlled pilot study to evaluate a technology platform for the assisted living of people with dementia and their carers. J Alzheimers Dis 2014;41(2):515-523. [doi: 10.3233/JAD-132156] [Medline: 24643137]
37. DuBenske LL, Gustafson DH, Namkoong K, Hawkins RP, Atwood AK, Brown RL, et al. CHESS improves cancer caregivers' burden and mood: results of an eHealth RCT. Health Psychol 2014 Oct;33(10):1261-1272 [FREE Full text] [doi: 10.1037/a0034216] [Medline: 24245838]
38. Kajiyama B, Thompson LW, Eto-Iwase T, Yamashita M, Di Mario J, Marian Tzuang Y, et al. Exploring the effectiveness of an internet-based program for reducing caregiver distress using the iCare Stress Management e-Training Program. Aging Ment Health 2013;17(5):544-554 [FREE Full text] [doi: 10.1080/13607863.2013.775641] [Medline: 23461355]
39. Kim J, Lee S, Kim J. Effects of a web-based stroke education program on recurrence prevention behaviors among stroke patients: a pilot study. Health Educ Res 2013 Jun;28(3):488-501. [doi: 10.1093/her/cyt044] [Medline: 23515115]
40. McLaughlin KA, Glang A, Beaver SV, Gau JM, Keen S. Web-based training in family advocacy. J Head Trauma Rehabil 2013;28(5):341-348 [FREE Full text] [doi: 10.1097/HTR.0b013e31824e1d43] [Medline: 22688211]
41. Smith GC, Egbert N, Dellman-Jenkins M, Nanna K, Palmieri PA. Reducing depression in stroke survivors and their informal caregivers: a randomized clinical trial of a Web-based intervention. Rehabil Psychol 2012 Aug;57(3):196-206 [FREE Full text] [doi: 10.1037/a0029587] [Medline: 22946607]
42. Pierce LL, Steiner VL, Khuder SA, Govoni AL, Horn LJ. The effect of a Web-based stroke intervention on carers' well-being and survivors' use of healthcare services. Disabil Rehabil 2009;31(20):1676-1684. [doi: 10.1080/09638280902751972] [Medline: 19479528]
43. Beauchamp N, Irvine AB, Seeley J, Johnson B. Worksite-based internet multimedia program for family caregivers of persons with dementia. Gerontologist 2005 Dec;45(6):793-801. [Medline: 16326661]
44. Pagán-Ortiz ME, Cortés DE, Rudloff N, Weitzman P, Levkoff S. Use of an online community to provide support to caregivers of people with dementia. J Gerontol Soc Work 2014;57(6-7):694-709 [FREE Full text] [doi: 10.1080/01634372.2014.901998] [Medline: 24689359]
45. Hattink BJJ, Meiland FJM, Overmars-Marx T, de Boer M, Ebben PWG, van Blanken M, et al. The electronic, personalizable Rosetta system for dementia care: exploring the user-friendliness, usefulness and impact. Disabil Rehabil Assist Technol 2016 Jul;11(1):61-71. [doi: 10.3109/17483107.2014.932022] [Medline: 24989993]
46. Gustafson DH, DuBenske LL, Namkoong K, Hawkins R, Chih M, Atwood AK, et al. An eHealth system supporting palliative care for patients with non-small cell lung cancer: a randomized trial. Cancer 2013 May 1;119(9):1744-1751 [FREE Full text] [doi: 10.1002/cncr.27939] [Medline: 23355273]
47. Steiner V, Pierce L. Building a web of support for caregivers of persons with stroke. Top Stroke Rehabil; 9(3) 2002:102-111 [FREE Full text] [doi: 10.1310/B9FE-QWEV-RM1F-FJA6]
48. Boutron I, Altman DG, Moher D, Schulz KF, Ravaud P, CONSORT NPT Group. CONSORT Statement for Randomized Trials of Nonpharmacologic Treatments: A 2017 Update and a CONSORT Extension for Nonpharmacologic Trial Abstracts. Ann Intern Med 2017 Jul 04;167(1):40-47. [doi: 10.7326/M17-0046] [Medline: 28630973]
49. Gilliam C, Steffen A. The relationship between caregiving self-efficacy and depressive symptoms in dementia family caregivers. Aging Ment Health; 10(2) 2006:79-86 [FREE Full text] [doi: 10.1080/13607860500310658]
50. Fortinsky R, Kercher K, Burant C. Measurement and correlates of family caregiver self-efficacy for managing dementia. Aging Ment Health; 6(2) 2002:153-160 [FREE Full text] [doi: 10.1080/13607860220126763]
51. Gallagher D, Ni MA, Crosby L, Ryan D, Lacey L, Coen R, et al. Self-efficacy for managing dementia may protect against burden and depression in Alzheimer's caregivers. Aging Ment Health; 15(6) 2011:663-670 [FREE Full text] [doi: 10.1080/13607863.2011.562179]
52. Au A, Lai M, Lau K, Pan P, Lam L, Thompson L, et al. Social support and well-being in dementia family caregivers: the mediating role of self-efficacy. Aging Ment Health; 13(5) 2009:761-768 [FREE Full text] [doi: 10.1080/13607860902918223]

53. 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]
54. Pingree S, Hawkins R, Baker T, duBenske L, Roberts LJ, Gustafson DH. The value of theory for enhancing and understanding e-health interventions. Am J Prev Med 2010 Jan;38(1):103-109 [FREE Full text] [doi: 10.1016/j.amepre.2009.09.035] [Medline: 20117565]
55. Canadian Institute for Health Research. 2017. Community-Based Primary Care URL: http://www.cihr-irsc.gc.ca/e/43626.html [accessed 2018-09-17] [WebCite Cache ID 72VXwOCm7]

Abbreviations

CCT: controlled clinical trial
GRADE: Grading of Recommendations Assessment, Development, and Evaluation system
RCT: randomized controlled trial
RoB: risk of bias
SMD: standardized mean difference

Edited by G Eysenbach; submitted 06.06.18; peer-reviewed by D Hansen, M Reblin; comments to author 18.07.18; revised version received 18.07.18; accepted 18.07.18; published 26.10.18.

Please cite as:
Ploeg J, Ali MU, Markle-Reid M, Valaitis R, Bartholomew A, Fitzpatrick-Lewis D, McAiney C, Sherifali D
Caregiver-Focused, Web-Based Interventions: Systematic Review and Meta-Analysis (Part 2)
J Med Internet Res 2018;20(10):e11247
URL: http://www.jmir.org/2018/10/e11247/
doi:10.2196/11247
PMID:30368439

Original Paper

Barriers and Facilitators When Implementing Web-Based Disease Monitoring and Management as a Substitution for Regular Outpatient Care in Pediatric Asthma: Qualitative Survey Study

Lara S van den Wijngaart[1], MSc, MD; Wytske W Geense[2], MSc; Annemie LM Boehmer[3], MD, PhD; Marianne L Brouwer[4], MD; Cindy AC Hugen[1], MD; Bart E van Ewijk[5,6], MD, PhD; Marie-José Koenen-Jacobs[7], MD; Anneke M Landstra[8], MD, PhD; Laetitia EM Niers[9], MD, PhD; Lonneke van Onzenoort-Bokken[9], MD; Mark D Ottink[10], MD; Eleonora RVM Rikkers-Mutsaerts[11], MD; Iris Groothuis[12], MD; Anja A Vaessen-Verberne[13], MD, PhD; Jolt Roukema[1], MD, PhD; Peter JFM Merkus[1], MD, PhD

[1]Department of Pediatric Pulmonology, Amalia Children's Hospital, Radboud Institute of Health Sciences, Radboud University Medical Center, Nijmegen, Netherlands

[2]IQ Healthcare, Radboud Institute of Health Sciences, Radboud University Medical Centre, Nijmegen, Netherlands

[3]Department of Pediatrics, Maasstad Hospital, Rotterdam, Netherlands

[4]Department of Pediatrics, Canisius Wilhelmina Hospital, Nijmegen, Netherlands

[5]Department of Pediatrics, Tergooi Hospital, Blaricum, Netherlands

[6]Department of Pediatrics, Tergooi Hospital, Hilversum, Netherlands

[7]Department of Pediatrics, Maas Hospital Pantein, Boxmeer, Netherlands

[8]Department of Pediatrics, Rijnstate Hospital, Arnhem, Netherlands

[9]Department of Pediatrics, Maxima Medical Center, Veldhoven, Netherlands

[10]Department of Pediatrics, Medical Spectrum Twente Hospital, Enschede, Netherlands

[11]Department of Pediatric Pulmonology, Leiden University Medical Center, Leiden, Netherlands

[12]Department of Pediatric Pulmonology, Juliana Children's Hospital, Haga Hospital, The Hague, Netherlands

[13]Department of Pediatrics, Amphia Hospital, Breda, Netherlands

Corresponding Author:
Lara S van den Wijngaart, MSc, MD
Department of Pediatric Pulmonology, Amalia Children's Hospital
Radboud Institute of Health Sciences
Radboud University Medical Center
Geert Grooteplein 10
Nijmegen, 6500 HB
Netherlands
Phone: 31 2414430
Fax: 31 2416428
Email: lara.vandenwijngaart@radboudumc.nl

Abstract

Background: Despite their potential benefits, many electronic health (eHealth) innovations evaluated in major studies fail to integrate into organizational routines, and the implementation of these innovations remains problematic.

Objective: The purpose of this study was to describe health care professionals' self-identified perceived barriers and facilitators for the implementation of a Web-based portal to monitor asthmatic children as a substitution for routine outpatient care. Also, we assessed patients' (or their parents) satisfaction with this eHealth innovation.

Methods: Between April and November 2015, we recruited 76 health care professionals (from 14 hospitals). During a period of 6 months, participants received 3 questionnaires to identify factors that facilitated or impeded the use of this eHealth innovation. Questionnaires for patients (or parents) were completed after the 6-month virtual asthma clinic (VAC) implementation period.

Results: Major perceived barriers included concerns about the lack of structural financial reimbursement for Web-based monitoring, lack of integration of this eHealth innovation with electronic medical records, the burden of Web-based portal use on clinician workload, and altered patient-professional relationship (due to fewer face-to-face contacts). Major perceived facilitators included enthusiastic and active initiators, a positive attitude of professionals toward eHealth, the possibility to tailor care to individual patients ("personalized eHealth"), easily deliverable care according to current guidelines using the VAC, and long-term profit and efficiency.

Conclusions: The implementation of Web-based disease monitoring and management in children is complex and dynamic and is influenced by multiple factors at the levels of the innovation itself, individual professionals, patients, social context, organizational context, and economic and political context. Understanding and defining the barriers and facilitators that influence the context is crucial for the successful implementation and sustainability of eHealth innovations.

(J Med Internet Res 2018;20(10):e284) doi:10.2196/jmir.9245

KEYWORDS

asthma; barriers and facilitators; eHealth; pediatric; Web-based monitoring

Introduction

In the last decades, the use of information and communication technology (ICT) in health care, also known as electronic health (eHealth), has rapidly evolved. It is widely used as a tool for improving health care, it delivers health care to geographically remote areas, it has the potential to reduce costs, and it improves health-related behavior and long-term management of chronic diseases. [1-3] Despite their potential benefits, many eHealth innovations evaluated in major studies do not reach the stage of structural implementation in daily practice and policy [4,5]. In The Netherlands, eHealth has no place yet in pediatric asthma management [6].

The barriers and facilitators for successful implementation have been the topic of several review papers [7-10], and Granja et al [7] suggested to also consider the scientific assessment of the quality of care and of the financial consequences of eHealth to optimize chances of successful implementation. We studied the implementation of eHealth prospectively, using our Web-based asthma monitoring management as an example.

In 2011, a Web-based portal to monitor asthmatic children was developed with the aim to change current pediatric asthma care and to (partly) substitute routine outpatient visits by Web-based monitoring. This Web-based portal, also called the virtual asthma clinic (VAC) for children, consisted of an information module, a forum for peers, a communication module, and an individual treatment plan for patients. Initially, we focused on providing evidence for a positive impact of this eHealth innovation on clinical endpoints. After 16 months of follow-up, the number of symptom-free days and the degree of asthma control improved in children who received care using the VAC, and we concluded that routine outpatient visits could partly be replaced by monitoring asthmatic children using eHealth. [11] Furthermore, the outcome of the cost-effectiveness analysis was in favor of using Web-based monitoring as a substitution for outpatient visits from a health economics perspective [12].

With this evidence of the impact of the VAC on clinical outcomes, the logical next step in the organizational process was the effective implementation of the VAC in a larger number of Dutch hospitals in order to become an integral part of current pediatric asthma management.

The main factor that complicates implementation is either human or organizational [13-16]. Often, the consequence is that eHealth is no longer supported when the funding supporting the study ends [17]. Additionally, future dissemination and implementation of the innovation are almost always neglected during its developmental stage, resulting in the poor sustainability of the innovation [17,18]. Thus, it is important to acquire a good understanding of the problem, the target group, its setting, the obstacles to change or resolve, and to "start with the end in mind." The first step in an implementation process should be to recognize, identify, and understand barriers and facilitators, crucial for addressing blockages to implementation and inventing strategies to improve the effective use and sustainability of eHealth in daily practice [19,20]. To examine barriers and facilitators, the model of Grol and Wensing can be used, proposing that barriers and facilitators can be examined at 6 different levels: innovation, individual professional, patient, social context, organizational context, and economic and political context (see Multimedia Appendix 1) [20].

Based on this model, we conducted a qualitative survey study with the aim to identify barriers and facilitators experienced by Dutch health care professionals and patients (or their parents) when implementing eHealth in routine pediatric asthma care. In addition, we summarize our lessons learned in recommendations that we consider relevant for successful implementation and sustainability of eHealth in general.

Methods

Study Design

This qualitative survey study utilized data collected from health care professionals including pediatricians, pediatric pulmonologists, nurse practitioners, pediatric (respiratory) nurses, and children or their parents [21]. We conducted the study in 14 hospitals (11 general and 3 tertiary, both urban and rural) in The Netherlands between April and December 2015.

Participants

The 8 hospitals originally participating in the randomized controlled trial (RCT) [11] were approached to also collaborate in this study. Of these, 6 hospitals agreed to participate, and 2 hospitals declined for logistic reasons. The other hospitals (8 in total) were contacted by email. All hospitals agreed to start with the implementation of the VAC. Finally, a total of 14 hospitals intended to participate in this study. In each hospital, several health care professionals confirmed their willingness to participate including pediatric pulmonologists, pediatricians, pediatric residents, nurse practitioners, or pediatric respiratory nurses. The number of participating health care professionals differed between hospitals. Also, patients (or their parents) were asked to complete a questionnaire about their satisfaction with the VAC after 6 months; they only completed 1 questionnaire at the end of the implementation (at 6 months).

Innovation

The VAC for children is a Web-based portal for children with asthma. It consists of a general information module and a secure private section where children (or their parents) can log in to communicate easily with their asthma team, download or consult their individual treatment plan, and complete a validated questionnaire for Web-based monitoring of their asthma control. Multimedia Appendix 2 shows different screenshots of the innovation (the individual care plan of a dummy patient).

Implementation of the VAC in pediatric asthma care included the use of the VAC in daily practice without a fixed protocol for the use of this innovation. Despite the existence of a national guideline for pediatric asthma management in The Netherlands, subtle differences between hospitals and health care professionals may exist (eg, provided information about asthma, the frequency of follow-up, preferences of prescribed medication). In addition, use of the VAC was tailored to local (organizational) preferences, for example, whether a nurse practitioner was involved in the use of the VAC or which professional was primarily responsible for contact with the patient. No restrictions were made about the frequency of outpatient visits or Web-based monitoring per patient.

Before implementation of the VAC, a 2-hour introductory visit was planned in each hospital, accessible to all participating professionals. These introductory visits were given by the same person (LvdW) and were aimed at adequately informing professionals about the purpose and details of the study and practical use of the VAC. The information provided was semistructured in each hospital. This visit included a real-time demonstration of the VAC and the different functions within the innovation. Further, questions by professionals were answered, and contact information and user manuals for patients and professionals were provided. Patients (or their parents) were informed about the use of the VAC by their asthma team and also received a written instruction manual. During the implementation process, all professionals could contact the help desk in case of existing problems.

Successful implementation was defined as using the VAC on a regular basis in daily pediatric asthma care, patients using the VAC regularly, and a positive attitude (of all end-users) toward continuing the use of the VAC in the future.

Data Collection and Analysis

Information about barriers and facilitators for successful implementation were collected by structured Web-based questionnaires (closed survey), which were completed by all professionals at the start of implementation and after 3 and 6 months using the same questionnaire at these different time points. Participation was voluntary, and health care professionals and patients could withdraw from the study at any time for any reason. No incentives were offered. Participants had 1 month to complete the questionnaire before the survey was closed.

Questionnaires were based on the model of Grol and Wensing, proposing that barriers and facilitators could be examined at 6 different levels: innovation, individual professional, patient, social context, organizational context, and economic and political context [20]. Questionnaires consisted of both open and multiple choice questions (see Multimedia Appendix 3). Participants received an email with a link to the questionnaire (email survey). No pilot-testing was done before distribution of these questionnaires, but the usability and technical functionality of the electronic questionnaire had been tested by the research team before fielding the questionnaire. All responses were automatically captured, analyzed, and anonymized.

In case of missing questionnaires, an automatically generated reminder was sent after 1 and 2 weeks. When a professional did not complete the questionnaire after 2 reminders, the link to the questionnaire was closed, and the questionnaire was considered missing. Only completed questionnaires were analyzed (as it was not possible to close the survey when items were missing).

Barriers and facilitators for implementation of the VAC in daily practice were identified from the open-ended questions and assessed using qualitative analysis. Barriers and facilitators were divided into different categories and organized into themes using the adapted model for understanding change at different levels of health care [20]. Two authors (LvdW and WG) identified and categorized the facilitators and barriers independently, based on the answers to the open-ended questions. In cases of disagreement, identification and categorization were discussed to reach consensus.

Descriptive statistics were used to analyze the multiple choice questions (Likert scale responses).

After 6 months, health care professionals were asked to send a questionnaire on patient satisfaction to their patients (or parents) via the VAC. The researchers had no influence on how many patients were reached because they had no access to local clinical data for medicolegal reasons. Therefore, no information was provided about the severity of their asthma and current medication or treatment regimes. Questions were addressed pertaining to the different modules of the VAC (Multimedia Appendix 4), the innovation, the (Children) Asthma Control Test questionnaire for Web-based monitoring disease deterioration, the communication module, the individual treatment plan, privacy and security, and overall satisfaction. Participants completed the questionnaire on a voluntary basis. Results of this questionnaire were used in this study and to update and further develop the VAC according to users' needs and wishes.

Ethics

The study was approved by the local medical ethics committee (Commisie Mensgebonden Onderzoek Nijmegen-Arnhem), which waived written informed consent.

Results

General Results

Health Care Professionals

The characteristics of all participating hospitals, including the number of participating health care professionals and whether implementation was successful or not, can be found in Multimedia Appendix 5.

Characteristics of the participating professionals, covering a wide range with respect to age and experience, are shown in Table 1. Initially, a total of 75 professionals intended to participate. The first questionnaire at the beginning of the study was completed by 68% (51/75) professionals: pediatric pulmonologists (14/51, 27%), pediatricians (19/51, 37%), nurse practitioners (7/51, 14%), and pediatric (respiratory) nurses (11/51, 22%). The response rates of the professionals at the beginning of the implementation and after 3 and 6 months were 67% (51/76), 63% (48/76), and 51% (39/76), respectively.

The mean age of all professionals was 47 (SD 8) years, and the mean work experience was 14 (SD 8) years. The majority of the professionals (37/51, 73%) had no experience with eHealth in daily practice at the start of the study.

Almost 70% (35/51) of the professionals were present at the introductory visit and received detailed information about the study and the use of the VAC. The other professionals were informed about the VAC by colleagues. Only 2% (1/51) did not receive any information before starting to use the VAC. All participants concluded that information provided at the introductory visits and the instruction manual were clear. Only 18% (9/51) of the professionals read the instruction manual thoroughly.

Implementation of the VAC was unsuccessful in 29% (4/14) hospitals. In 1 hospital, this was due to insufficient staff; just 1 pediatric pulmonologist was available for the implementation of the VAC, and for this physician, the time investment was too much in addition to his regular work. No specific reasons for unsuccessful implementation in the other 3 hospitals could be assessed due to missing questionnaires at follow-up. No further follow-up was done with the hospital representatives who initially agreed to participate.

Characteristics of Patients and Their Parents

At the end of the study, 66 parents completed the voluntary questionnaire to provide information about the VAC. Most of their children were male (51/66, 77%) with a mean age of 10.1 (SD 2.5) years. On average, these parents used the VAC for 7.6 (SD 4.7) months. There were 10 children with asthma who completed the questionnaire.

Barriers and Facilitators

Barriers and facilitators were divided into 6 categories and organized into themes using the adapted model for understanding change at different levels of health care [20]. Table 2 provides an overview of the framework with barriers and facilitators in each theme.

Innovation

By *innovation*, we mean “the eHealth innovation (or the object) of the implementation process,” in this study, the VAC. There were 4 categories related to this theme: (1) attractiveness; (2) amount of information; (3) (dis)advantage; and (4) accessibility and usability.

Several professionals stated the *attractiveness* of the innovation as a facilitator, as the innovation was clear, easy, and user-friendly. Also, the innovation provided bundled, reliable, and age-adjusted *information* about asthma for children and their parents.

Professionals also mentioned the *advantage* of the VAC to substitute routine outpatient visits as a major facilitator. However, other professionals defined restrictions in the use of the innovation. For example, their assumption or cognition was that Web-based monitoring only based on the asthma control test, which is sensitive to confounding factors, is not an adequate substitution for these visits. They also stated that the missing integration between the VAC and electronic medical records (EMRs) was a barrier for the use of the VAC in daily practice, as this resulted in an extra workload to keep both systems up-to-date.

Interactive educational methods (ie, demonstration and instruction of the VAC at the beginning of the study), enthusiasm, and motivation of those responsible for the innovation were considered important contributors to the *accessibility* of the innovation. The accessibility of the innovation was also facilitated by the available user manual and (technical) support service during the study. Both items were central facilitators in helping professionals to use the eHealth innovation in daily practice. Other professionals had some concerns about the privacy of patient data when using eHealth and concerns about the *usability* of the innovation when ICT problems occurred.

Table 1. Characteristics of participating health care professionals.

Characteristics	Pediatric pulmonologist (n=14)	Pediatrician (n=19)	Nurse practitioner (n=7)	Pediatric (respiratory) nurse (n=11)
Age (years), mean (SD)	47 (8.5)	46 (9.6)	50 (5.1)	48 (5.7)
Gender (male), n (%)	7 (50)	9 (47)	0 (0)	0 (0)
Work experience (years), mean (SD)	10 (6.5)	12 (8.0)	18 (9.1)	18 (6.0)
Present at introductory visit, n (%)	9 (64)	13 (68)	5 (71)	8 (73)

Table 2. Themes, categories, and facilitators and barriers.

Themes and category	Facilitators	Barriers
Innovation		
Attractiveness	• User-friendliness of the program	• Usage restriction of the innovation
Amount of information	• Bundled, reliable, and age-adjusted information about asthma	• Asthma control test is sensitive but not specific for poor asthma control (eg, a common cold may falsely suggest poor asthma control)
(Dis)advantage	• Possibility of frequent monitoring patients' symptoms	• Link with patients' electronic health records is missing
Accessibility and usability	• Possibility to ask questions (helpdesk) • User manual with instruction at start	• The (theoretical) possibility that privacy of patients' data could not be fully guaranteed • Usability in case of ICT[a] problems
Individual professionals		
Attitude of professionals	• Attitude of the health care professional (believe in eHealth, convinced of the value of the innovation in daily practice)	• Attitude of the health care professional (not convinced of the value of the innovation in daily practice)
Professional skills	• Enrichment of work	• No substitution for face-to-face contact
Knowledge and awareness of eHealth	• Expectation that parents and children favor the innovation • Accessible contact and improvements in relationship with patients • More time for complex patients • Customizing care to the individual patient • Possibility to gain experience with eHealth • Results of the randomized controlled trial about effectiveness of the innovation were positive • Experience with the use of the innovation in an earlier study	• Risk of losing patients out of sight • Extra way of communication • Adequate and timely response to messages is difficult to ensure • Less patient contact can have a negative effect on the professional's own development • Difficulty to motivate colleagues to use the innovation • Lack of time to explore the innovation • Management imposed the innovation • Difficulty to recruit patients • No routine use of the innovation • Lack of knowledge or (computer) skills
Patients		
(Dis)advantage for the patient	• Fewer outpatient visits	• Less contact with the health care professional
Patient satisfaction and compliance	• Less absenteeism from school	• More (daily) confrontation with the diagnosis of asthma
Accessibility and usability	• Promoting patients' compliance, self-management, and knowledge • Patient satisfaction • Less focus on illness of the child • Improvement of security and privacy with the use of the innovation	• The innovation is not applicable for all patients • Ensure continuous use of the innovation by patients • Lack of access to the internet • Different options in the innovation • Patients were inadequately instructed about the innovation
Social context		
(Lack of) sufficient interprofessional collaboration	• Better collaboration between health care professionals	• Unclear allocation of tasks between health care professionals
Substitution of tasks or care by health care professionals	• Substitution of care by health care professionals	—[b]
Care according to current guidelines	• Possibility to give care according to the most recent guidelines	—[b]
Organizational context		
Organization of care or care processes	• Care logistics is better organized nowadays	• Implementation of the innovation in daily practice was unclear or incomplete

Themes and category	Facilitators	Barriers
Time	- Fewer outpatient visits resulting in more time	- Link to patients' electronic health dossiers is missing
ICT infrastructure	- Time saving and efficient - Smaller workload for personnel - Positive public relations for department and hospital	- Lack of involvement of management during implementation - Lack of promotional material (ie, leaflets) - Extra workload, time, and administration - Delay at start-up - ICT problems
Economic and political context		
Financial arrangements	- Complementary to current asthma management - Keep up with current developments - Improvement of care	- Uncertainty of future of pediatricians - Similar eHealth innovations had no added values - Lack of reimbursement for Web-based monitoring

[a] ICT: information and communication technology.

[b] Barriers were not described in this theme and/or category.

Individual Professionals

There were 3 categories defined: (1) knowledge and awareness of eHealth; (2) attitude of health care professionals toward the innovation; and (3) professional skills.

At the start of the study, 20% (10/51) of the professionals concluded they had insufficient *knowledge* about the VAC and for whom it was intended. Almost 40% (20/51) did not know what was expected with regard to working with the VAC, and 8% (4/51) had doubts about its use in daily practice.

The professionals' commitment toward eHealth appeared to be of great importance for the successful implementation of an eHealth innovation. A *positive attitude* toward the use of eHealth seemed the most important facilitator for easy and quick implementation of the VAC in daily practice. By using the VAC, professionals experienced eHealth as an enrichment of their work. They were more able to customize care to the individual patient, and this "personalized eHealth" matches the needs and wishes of both professionals and patients. Besides this, the use of the VAC was reported to facilitate the provision of asthma care according to current guidelines. Also, they emphasized the accessible contact, the improvement of the patient-doctor relationship, and the fact that they had more time for complex patients.

In contrast, when professionals *were not convinced* of the added value of the eHealth innovation, successful implementation of the innovation was unlikely despite all efforts. As reasons, professionals mentioned the risk of losing sight of patients and the fact that Web-based monitoring could never substitute face-to-face contacts due to, for example, the missing interpersonal relationship. Additionally, other professionals saw the VAC as a possible barrier for professionals' own development due to fewer patient contacts.

Nevertheless, most professionals had a positive attitude about eHealth overall (*"it is part of health care"*). They stated that with the use of the VAC they had a great opportunity to start using eHealth in daily practice in a structured, solid, and easy way.

Professional skills and the cognition of professionals toward these skills were also important. Some professionals experienced a lack of (eHealth, ICT, or computer) knowledge or a lack of (computer) skills. Further, they stated that there was no time to explore the VAC adequately and improve their skills. The majority of the professionals mentioned that they wanted to spend more time using the VAC to continue to improve these skills in order to take full advantage of all the various features of the innovation. However, they lacked time due to their work and the responsibilities of taking care of patients.

Patient Level

At the level of the patient, professionals mentioned several facilitators and barriers for the implementation of the VAC, which were divided into 3 categories: (1) (dis)advantage for the patient; (2) patient satisfaction and compliance; and (3) accessibility and usability.

The *main advantages* reported by patients were fewer routine outpatient visits and, as a result of this, less absence from school and work (parent-related). Professionals experienced high *patient satisfaction*. Further, they stated that *patients' compliance*, self-management, and knowledge were promoted by the use of the VAC. With the use of the VAC, patients can easily and proactively participate in the treatment of their own disease, and asthma management is more based on shared decision making. Also, Web-based monitoring resulted in less focus on the illness of the child.

Notably, for other professionals, this last argument was not a facilitator but a barrier. They mentioned that Web-based monitoring emphasized the confrontation with asthma, as a result of the monthly questionnaire. The fact that the VAC is not suitable for all patients was found to be another barrier to implementation according to professionals. For example, patients with poor perception of their asthma symptoms were not considered suitable for Web-based monitoring as a substitution for routine outpatient visits. Therefore, they emphasized that less contact with professionals is not always feasible or wise. A potential risk of Web-based monitoring via the VAC is the patient compliance: how can professionals ensure continuous use of the innovation by patients? Compliance may be influenced by patients' motivation, symptom burden, and cognitive abilities. With a patient's lack of motivation for completing the assessment tool, Web-based monitoring as a substitution for outpatient visits is impossible and will not result in proper asthma care. Another potential barrier mentioned by professionals was the concern about a lack of access to the internet for some patients, resulting in selection bias and the innovation not being *accessible* for all patients. However, in The Netherlands, over 96% of the Dutch households had access to the internet in 2016.

The majority (56/60, 93.3%) of the parents concluded that the VAC had added value to current daily pediatric asthma management. Parents were satisfied with the attractiveness, usefulness, and user-friendliness of the innovation. Further, the C-ACT questionnaire was found to be an adequate tool for Web-based monitoring of disease deterioration. Because of the frequent follow-up by C-ACT, most parents were more aware of their child's asthma symptoms. They concluded that this eventually led to better asthma control. The perceived threshold to communicate through the VAC was low, communication was easy, and they concluded that the asthma team responded within the agreed timeframe (of 2 working days) most of the time. The majority of the parents experienced more control of the asthma of their child, felt to be more independent, and experienced better cooperation with the physician by using the VAC. Security was no issue for 96% (58/60) of the patients or parents, as they trusted that personal information on the VAC was safe. Only 10% (6/60) of the parents missed out on the personal or face-to-face contact with professionals. Overall, the majority of the parents (56/60, 93%) concluded that the VAC had additional value to current daily pediatric asthma management and wanted to use the VAC in the future. The VAC was scored an 8 or higher (based on a 10-point scale, with 10 corresponding to maximum satisfaction) by 78% (47/60) of the parents. Similar scores were reported by the children completing the questionnaire.

Social Context

There were 3 categories defined at the level of social context: (1) (lack of) sufficient interprofessional collaboration; (2) substitution of tasks or care by health care professionals; and (3) care according to current guidelines.

First, regarding *sufficient interprofessional collaboration*, most professionals stated that the use of an eHealth innovation like the VAC improved the collaboration between professionals. However, this was only the case when all participating professionals adjusted their tasks before starting to use the VAC. When the asthma team did not adjust or coordinate tasks accordingly, professionals experienced a lack of sufficient interprofessional collaboration resulting in inadequate use of the eHealth innovation.

Second, the use of the VAC facilitated *substitution of tasks* between professionals. For example, nurse practitioners completed the individual care plan for all patients instead of the pediatrician. The latter reported to spend more time on more complex patients.

Third, a major facilitator for the implementation of the VAC was the fact that this eHealth innovation makes it possible for professionals to easily optimize and individualize care for the patients to *current guidelines*. The layout of the individual care plan was based on the content of the national pediatric asthma guideline, including treatment goals, medication, information, and an action plan in case of an asthma exacerbation.

Organizational Context

There were 3 categories related to this theme: (1) organization of care processes or organizational structure; (2) time; and (3) ICT infrastructure.

A better *organization of care* by the use of VAC was considered an important driving factor for successful implementation. For example, the easy and accessible way of communicating with patients via the VAC gave professionals the possibility to contact the patients at a time during the day that was beneficial for them. Despite some professionals experiencing *lower workload* through the use of the VAC, others emphasized the *extra workload and time* due to the extra administration for completing all information in the individual treatment plans.

Contrastingly, the involvement of the medical management at an early stage of the implementation made it easier for professionals to change work processes and, at the beginning of the implementation, spend *more time* on the use of the eHealth innovation.

A major barrier was mentioned by almost all health care professionals, namely *the lack of interoperability* (which is defined as the ability of a system to work with other systems without special effort on the part of the customer) between the VAC and current EMRs. For medicolegal reasons, professionals have to include all patients' contacts in their personal EMR. Due to the missing link between the VAC and current EMRs, professionals had to complete both portals, which was time-consuming and not workable in daily practice. A sustainable integration is necessary to keep both systems up-to-date easily. Local ICT problems caused minor barriers, for example, firewall blocking the website.

Economic and Political Context

Only 1 category was defined in this theme: financial arrangements. The most important barrier for successful implementation of the VAC was the *lack of structural financial reimbursement* for Web-based monitoring. When Web-based monitoring substituted 50% of the routine outpatient visits, professionals were only reimbursed for 50% of these visits. However, they had to invest time in the use of the VAC (eg, completing the individual treatment plans, respond to messages etc) and were encouraged to reduce the number of routine outpatient visits.

Discussion

Principal Findings

Major perceived barriers found in this study included concerns about lack of interoperability of this innovation with other systems, the lack of structural financial reimbursement for Web-based monitoring, the burden of eHealth use on professionals' workload, and a changing patient-professional relationship (due to less face-to-face contact).

Major perceived facilitators included training and support, a positive attitude of professionals toward eHealth, the advantages for the patients, and the possibility to tailor care to the individual patient ("personalized eHealth"), to substitute tasks between professionals, and to easily deliver care according to current guidelines by using the VAC.

We identified these barriers and facilitators for the implementation of a Web-based portal to monitor asthmatic children as a substitution for routine outpatient visits among a group of health care professionals, patients, and parents in 14 hospitals. These barriers and facilitators were arranged in 6 themes as described by Grol and Wensing: the innovation itself, individual professionals, patients, social context, organizational context, and political and economic context [20]. Although barriers and facilitators differ in scope, context, and strength, most of them were shared by professionals from different hospitals and with different functions. Findings of this study suggest and confirm that issues around implementation are multilevel and complex, and it is important to mention that no single factor was identified as a key barrier or facilitator.

The *lack of interoperability* is perhaps one of the greatest organizational obstacles to the long-term integration of eHealth infrastructure in health care [16,22,23]. As with most of the existing eHealth networks, the VAC tends to rely on custom-built systems made for specific users operating in specific settings, has a lack of open connectivity with other systems (eg, EMRs), and no information is exchanged automatically. This results in higher than necessary workload for professionals since the EMR has to contain the latest and relevant medical information due to medicolegal reasons. An adequate interface with other ICT systems is essential for acceptance, use, successful implementation, and long-term sustainability of the VAC.

The other significant obstacle for further diffusion of our eHealth innovation is the *lack of structural financial reimbursement* of Web-based monitoring [14,24]. To date, there is still a lack of payment for electronically delivered care, and funding of the current health system is solely based on reimbursement of face-to-face contacts. Using the VAC as a (partial) substitution for outpatient visits lead to production loss for hospitals and less reimbursement of these outpatient visits. Nevertheless, instead of seeing patients at the outpatient clinic, professionals now have to invest time in using the VAC (eg, completing the individual treatment plans, respond to messages, etc) for the Web-based monitoring of patients. This type of care was not yet reimbursed, and the incentive for professionals to use the VAC was solely based on their goodwill and grants. Logically, this is not sustainable and adequate funding for Web-based monitoring as a substitution for routine care, in combination with supportive policy and political vision toward eHealth, is crucial to acheiving long-term goals.

eHealth innovations have to fit the workflow of professionals to minimize *the burden of eHealth use*. Otherwise, they experience an extra workload and a time constraint, which are major impediments to implementation [14,16,25,26]. The reason for an extra workload with the introduction of the VAC consisted mostly of a lack of time to learn how to use the VAC, the work involved in transferring information between 2 systems, and an alteration in established professional roles, responsibilities, and work styles. Providing a transition period in which users become familiar with and learn how to use the eHealth innovation has been desirable to facilitate implementation [16,27].

Limitations and Strengths

The main methodological limitation of this study is that the data were collected using 3 consecutive questionnaires over a period of 6 months. In-depth interviews to assess more detailed information about the barriers and facilitators perceived by professionals were not used. Further, we had developed a questionnaire based on the model of Grol and Wensing, which was not validated and tested for reliability. However, no valid questionnaire to assess barriers and facilitators for implementation has been published or validated. In addition, we were not able to specify similarities and differences between respondents and nonrespondents as we did not have any data of the latter. Therefore, we cannot rule out some bias. Nevertheless, we consider the response rate (51/76, 67%) as quite acceptable and therefore assume that a bias may be minimal.

For the collection of parents' and patients' data, we were dependent on health care professionals sending this questionnaire via the VAC because we were not allowed to have access to patients' data. This may have influenced the response rate of this questionnaire, as it is possible that professionals did not send this questionnaire or sent it too late. Also, it would be interesting to compare the responses of the health care professionals and patients or parents in future research since their perceptions or viewpoints may be different. Another limitation could be the fact that all 3 questionnaires for professionals were sent in a time period of 6 months after starting to use the VAC. This means that patients monitored over the Web using the VAC were only seen once during this period. This could make it difficult for professionals to develop an objectively based opinion about the importance of the VAC for their daily practice. We could not objectively assess possible changes in time, differences between hospitals, or type of professionals. Also, it is possible that the time period of the study (6 months) was too short to identify all key barriers and facilitators. Nevertheless, most findings were in line with earlier studies.

An overall strength of the VAC is the involvement of health care professionals and patients from the beginning of the project. Participants' wishes and possible objections against the VAC were used for its development, and after the RCT their feedback was used to update the VAC to optimize their use before wider implementation. Ongoing involvement of key stakeholders (especially end-users) in the design and development of eHealth technologies helps to ensure that systems are likely to be valued and used by professionals (intrinsic motivation) and should be considered as a way of overcoming barriers and facilitating implementation [16,28-31].

Textbox 1. Recommendations for successful implementation of an eHealth innovation.

- Quality of care (eg, efficacy, superiority) is a major facilitator for the implementation of an eHealth innovation and appears to be relevant for the adoption of eHealth. Therefore, this should be assessed before implementation of the innovation in daily practice.
- Besides health care professionals and patients, other stakeholders (eg, information and communication technology services, governments, insurance providers) should be included as early as possible in different development and implementation phases.
- Involve end-users in the design and development of an eHealth innovation to make them develop feelings of ownership toward the innovation and to achieve a positive perception of the innovation's usefulness and user-friendliness. Use their feedback to optimize the innovation.
- Maximize the use of and adherence to eHealth applications by developing and modifying the eHealth app from the patient's point of view: only when a clear advantage or incentive is perceived, eHealth may sustain. The lower frequency of outpatient clinic visits is probably a major incentive in the case of the virtual asthma clinic.
- Provide adequate training and support (eg, technical assistance, training on-the-job and user manuals) during and after the implementation process to optimize use of the innovation.
- Aim for a transition period for users. This period can be used to become familiar with and learn how to use the eHealth innovation and thereby facilitate the implementation process.
- Even the best eHealth system will fail if users do not have the right knowledge and skills to use it effectively.
- Standards for technology should be optimized before or during the implementation process (interoperability, security, and privacy of hospital information and communication technology systems).
- Sufficient financial reimbursement needs to be arranged to support sustainable implementation. Chances of finding reimbursement increase with evidence of efficacy, superiority, and cost-effectiveness.
- Guarantee the privacy and safety of patients' data.
- Implementing eHealth and sustaining it are very different ventures. After implementation, ongoing monitoring, evaluation, and adoption of the system are needed to ensure intended goals are being met and to identify ongoing barriers.

Another strength of this implementation study is that the questionnaires were based on factors identified by qualitative research, the model of Grol and Wensing [20], and thereby represent the complete spectrum of important factors related to acceptance and diffusion of an eHealth innovation. Another strong point is the fact that we obtained a representative sample of participants. Professionals from different hospitals, both secondary and tertiary care, participated in the study. Hospitals were located in both rural and urban areas. Further, professionals had different functions in the asthma team (eg, pediatric pulmonologist, pediatricians, nurse practitioners, etc). This gives a good reflection of pediatric asthma care.

Based on the barriers and facilitators found in this study, we propose a list of recommendations when implementing eHealth innovations (Textbox 1). These can be used to tailor organizational improvements to the needs of individuals and organizations.

Therefore, the suitability of the innovation in its context is essential for an innovation to become successful. Professionals can be seen and approached as key holders to achieve change and successful implementation of the innovation [7]. However, management support is essential to encourage professionals to hold on to their innovative behavior, to embrace new strategies for improving health care, to ensure sustainable use of the innovation, and to implement renewal. As mentioned, patients must also be involved in the innovation as cocreators to make them develop feelings of ownership toward the innovation and enhance positive perception of the innovation's usefulness and user-friendliness. Only when a clear advantage or incentive is perceived by patients, are they inclined to continue to adhere to and use eHealth and, thus, it may sustain. The lower frequency of outpatient clinic visits is probably a major incentive in the case of the VAC. Besides an organizational change within the hospital, other structural changes are also necessary to achieve successful implementation of eHealth. Other stakeholders, such as the Government, policy makers, and insurance companies, have a crucial role in the implementation of eHealth, nowadays and in the future. Without adequate financial reimbursement, the introduction of innovations in health care has been doomed to fail. Sustainable solutions to finance new strategies are necessary to ensure large-scale innovations to conquer a solid place within the current health care system. At an early stage, all stakeholders together have to assess which innovations have the potential to change health care positively, from the perspective of value-based health care, efficiency, and cost savings.

Conclusion

Implementation of an eHealth innovation is a complex, dynamic process influenced by multiple factors at the levels of the innovation itself, individual professionals, patients, and the social, organizational, and economic and political context. Understanding and defining the barriers and facilitators that influence the context appears to be important for the successful implementation and sustainability of an innovation.

Despite serious concerns about structural financial reimbursement and the lack of interoperability of the system, health care professionals, patients, and parents generally expressed enthusiasm about the VAC in our study. In the majority of the participating hospitals, this innovation became part of regular asthma care for children and is it used as a (partial) substitution of traditional outpatient care for children with asthma.

Acknowledgments

We gratefully thank all collaborating pediatricians, nurse practitioners, and pediatric (respiratory) nurses as well as patients (and their parents) for their contribution to this study, their enthusiasm, and their efforts in using this eHealth innovation in daily practice.

The RCT study was funded by Lung Foundation Netherlands (Grant No. 3.4.10.016) and Dutch Innovation Foundation Health insurance companies (Grant No. B 10-135). This implementation study was funded by ZonMw (Project No. 430000012). The funders had no role in the design of the studies, data collection, analysis, or reporting.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Barriers and facilitators for change at different levels of health care according to Model of Grol and Wensing.

[PDF File (Adobe PDF File), 173KB - jmir_v20i10e284_app1.pdf]

Multimedia Appendix 2

Private section of the virtual asthma clinic.

[PDF File (Adobe PDF File), 1011KB - jmir_v20i10e284_app2.pdf]

Multimedia Appendix 3

Questionnaire for health care professionals.

[PDF File (Adobe PDF File), 193KB - jmir_v20i10e284_app3.pdf]

Multimedia Appendix 4

Questionnaire for patients and parents.

[PDF File (Adobe PDF File), 193KB - jmir_v20i10e284_app4.pdf]

Multimedia Appendix 5

Characteristics of participating hospitals.

[PDF File (Adobe PDF File), 177KB - jmir_v20i10e284_app5.pdf]

References

1. McLean S, Chandler D, Nurmatov U, Liu J, Pagliari C, Car J, et al. Telehealthcare for asthma: a Cochrane review. Canadian Medical Association journal 2011;183(11): [FREE Full text] [doi: 10.1503/cmaj.101146] [Medline: 21746825]
2. Zhao J, Zhai Y, Zhu W, Sun D. Effectiveness of Telemedicine for Controlling Asthma Symptoms: A Systematic Review and Meta-analysis. Telemed J E Health 2015 Jun;21(6):484-492. [doi: 10.1089/tmj.2014.0119] [Medline: 25393915]
3. Ekeland A, Bowes A, Flottorp S. Effectiveness of telemedicine: a systematic review of reviews. Int J Med Inform 2010 Nov;79(11):736-771. [doi: 10.1016/j.ijmedinf.2010.08.006] [Medline: 20884286]
4. Jung M, Berthon P. Fulfilling the promise: a model for delivering successful online health care. J Med Marketing 2009;9(3):. [doi: 10.1057/jmm.2009.26]
5. Shortliffe EH. Strategic action in health information technology: why the obvious has taken so long. Health Aff (Millwood) 2005;24(5):1222-1233. [doi: 10.1377/hlthaff.24.5.1222] [Medline: 16162567]
6. Pijnenburg MW, Baraldi E, Brand PLP, Carlsen K, Eber E, Frischer T, et al. Monitoring asthma in children. Eur Respir J 2015 Apr;45(4):906-925 [FREE Full text] [doi: 10.1183/09031936.00088814] [Medline: 25745042]
7. Granja C, Janssen W, Johansen MA. Factors Determining the Success and Failure of eHealth Interventions: Systematic Review of the Literature. J Med Internet Res 2018 May 01;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
8. Banbury A, Nancarrow S, Dart J, Gray L, Parkinson L. Telehealth Interventions Delivering Home-based Support Group Videoconferencing: Systematic Review. J Med Internet Res 2018 Feb 02;20(2):e25 [FREE Full text] [doi: 10.2196/jmir.8090] [Medline: 29396387]

9. Slater H, Campbell JM, Stinson JN, Burley MM, Briggs AM. End User and Implementer Experiences of mHealth Technologies for Noncommunicable Chronic Disease Management in Young Adults: Systematic Review. J Med Internet Res 2017 Dec 12;19(12):e406 [FREE Full text] [doi: 10.2196/jmir.8888] [Medline: 29233804]
10. Matthew-Maich N, Harris L, Ploeg J, Markle-Reid M, Valaitis R, Ibrahim S, et al. Designing, Implementing, and Evaluating Mobile Health Technologies for Managing Chronic Conditions in Older Adults: A Scoping Review. JMIR Mhealth Uhealth 2016;4(2) [FREE Full text] [doi: 10.2196/mhealth.5127] [Medline: 27282195]
11. van den Wijngaart LS, Roukema J, Boehmer A, Brouwer ML, Hugen CAC, Niers LEM, et al. A virtual asthma clinic for children: fewer routine outpatient visits, same asthma control. European Respiratory Journal 2017 Oct 5;50(4). [doi: 10.1183/13993003.00471-2017] [Medline: 28982775]
12. van den Wijngaart LS, Kievit W, Roukema J, Boehmer ALM, Brouwer ML, Hugen CAC, et al. Online asthma management for children is cost-effective. European Respiratory Journal 2017 Oct 5;50(4). [doi: 10.1183/13993003.01413-2017] [Medline: 28982768]
13. Mair F, May C, O'Donnell C, Finch T, Sullivan F, Murray E. Factors that promote or inhibit the implementation of e-health systems: an explanatory systematic review. Bull World Health Organ 2012;90(5) [FREE Full text]
14. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implementation science : IS 2016;11(1) [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
15. Murray E, Burns J, May C, Finch T, O'Donnell C, Wallace P, et al. Why is it difficult to implement e-health initiatives? A qualitative study. Implement Sci 2011 Jan 19;6:6 [FREE Full text] [doi: 10.1186/1748-5908-6-6] [Medline: 21244714]
16. Gagnon M, Desmartis M, Labrecque M, Car J, Pagliari C, Pluye P, et al. Systematic review of factors influencing the adoption of information and communication technologies by healthcare professionals. J Med Syst 2012 Feb;36(1):241-277 [FREE Full text] [doi: 10.1007/s10916-010-9473-4] [Medline: 20703721]
17. Glasgow RE, Phillips SM, Sanchez MA. Implementation science approaches for integrating eHealth research into practice and policy. Int J Med Inform 2014 Jul;83(7):e1-11. [doi: 10.1016/j.ijmedinf.2013.07.002] [Medline: 23910896]
18. Klesges L, Estabrooks P, Dzewaltowski D, Bull S, Glasgow R. Beginning with the application in mind: designing and planning health behavior change interventions to enhance dissemination. Annals of behavioral medicine : a publication of the Society of Behavioral Medicine. Suppl 2005;29:a. [doi: 10.1207/s15324796abm2902s_10] [Medline: 15921491]
19. Grol R, Grimshaw J. From best evidence to best practiceffective implementation of change in patients' care. Lancet 2003;362(9391):e [FREE Full text] [doi: 10.1016/S0140-6736(03)14546-1]
20. Grol R, Wensing M. What drives change? Barriers to and incentives for achieving evidence-based practice. Med J Aust 2004;180(6 Suppl):S57-S60 [FREE Full text] [Medline: 15012583]
21. Jansen H. The Logic of Qualitative Survey Research and its Position in the Field of Social Research Methods. Forum Qualitative Sozialforschung/Forum: Qualitative Social Research 2010;11(2) [FREE Full text] [doi: http://dx.doi.org/10.17169/fqs-11.2.1450]
22. Miller EA. Solving the disjuncture between research and practice: telehealth trends in the 21st century. Health Policy 2007 Jul;82(2):133-141. [doi: 10.1016/j.healthpol.2006.09.011] [Medline: 17046097]

. Craft RL. Toward technical interoperability in telemedicine. Telemedicine journal and e-health : the official journal of the American Telemedicine Association 2005;11(3):384-404. [doi: 10.1089/tmj.2005.11.384] [Medline: 16035933]

24. Police RL, Foster T, Wong KS. Adoption and use of health information technology in physician practice organisations: systematic review. Inform Prim Care 2010;18(4):245-258 [FREE Full text] [Medline: 22040851]
25. Tieu L, Sarkar U, Schillinger D, Ralston J, Ratanawongsa N, Pasick R, et al. Barriers and Facilitators to Online Portal Use Among Patients and Caregivers in a Safety Net Health Care System: A Qualitative Study. Journal of medical Internet research 2015;17(12):A [FREE Full text] [doi: 10.2196/jmir.4847] [Medline: 26681155]
26. Lluch M. Healthcare professionals' organisational barriers to health information technologies-a literature review. Int J Med Inform 2011 Dec;80(12):849-862. [doi: 10.1016/j.ijmedinf.2011.09.005] [Medline: 22000677]
27. Rahimi B, Vimarlund V, Timpka T. Health information system implementation: a qualitative meta-analysis. J Med Syst 2009;33(5):a [FREE Full text]
28. Goldstein D, Phelan R, Wilson R, Ross-White A, VanDenKerkhof E, Penning J, et al. Brief review: Adoption of electronic medical records to enhance acute pain management. Can J Anaesth 2014;61(2):164-179 [FREE Full text] [doi: 10.1007/s12630-013-0069-6]
29. Cresswell K, Sheikh A. Organizational issues in the implementation and adoption of health information technology innovations: an interpretative review. International journal of medical informatics 2013;82(5):e73-e86. [doi: 10.1016/j.ijmedinf.2012.10.007] [Medline: 23146626]
30. Paré G, Sicotte C, Jacques H. The effects of creating psychological ownership on physicians' acceptance of clinical information systems. J Am Med Inform Assoc 2006;13(2):197-205 [FREE Full text] [doi: 10.1197/jamia.M1930] [Medline: 16357351]
31. Swinkels ICS, Huygens MWJ, Schoenmakers TM, Oude NW, van Velsen L, Vermeulen J, et al. Lessons Learned From a Living Lab on the Broad Adoption of eHealth in Primary Health Care. J Med Internet Res 2018 Mar 29;20(3):e83 [FREE Full text] [doi: 10.2196/jmir.9110] [Medline: 29599108]

Abbreviations

C-ACT: Childhood Asthma Control Test
eHealth: electronic health
EMR: electronic medical record
ICT: information and communication technology
RCT: randomized controlled trial
VAC: virtual asthma clinic

Edited by G Eysenbach; submitted 07.11.17; peer-reviewed by P Pluye, R El Sherif, R Estrada; comments to author 29.01.18; revised version received 17.05.18; accepted 21.06.18; published 30.10.18.

Please cite as:
van den Wijngaart LS, Geense WW, Boehmer ALM, Brouwer ML, Hugen CAC, van Ewijk BE, Koenen-Jacobs MJ, Landstra AM, Niers LEM, van Onzenoort-Bokken L, Ottink MD, Rikkers-Mutsaerts ERVM, Groothuis I, Vaessen-Verberne AA, Roukema J, Merkus PJFM
Barriers and Facilitators When Implementing Web-Based Disease Monitoring and Management as a Substitution for Regular Outpatient Care in Pediatric Asthma: Qualitative Survey Study
J Med Internet Res 2018;20(10):e284
URL: http://www.jmir.org/2018/10/e284/
doi:10.2196/jmir.9245
PMID:30377147

Original Paper

Web-Based Health Intervention for Young People Who Have a Parent with a Mental Illness: Delphi Study Among Potential Future Users

Jodie L Matar[1*], GDipPsych (Hons); Darryl J Maybery[2*], PhD; Louise A McLean[3*], PhD; Andrea Reupert[3*], PhD

[1]Faculty of Education, Monash University, Clayton, Australia
[2]Department of Rural Health, Monash University, Moe, Australia
[3]Education, Psychology Programs, Monash University, Clayton, Australia
[*]all authors contributed equally

Corresponding Author:
Andrea Reupert, PhD
Education, Psychology Programs
Monash University
19 Ancora Imparo Way
Clayton, 3800
Australia
Phone: 61 399024587
Email: andrea.reupert@monash.edu

Abstract

Background: Young people who have a parent with a mental illness face elevated risks to their mental health and well-being. However, they may not have access to appropriate interventions. Web-based interventions may reach and meet the needs of this at-risk group, yet their preferences regarding the features of this medium are unknown.

Objective: This study sought to determine the utility of a Web-based intervention to meet the needs of young people who have a parent with a mental illness and their perspectives regarding the types of features of such a website.

Methods: A systematic, 2-round Delphi study was employed to solicit the views of 282 young people aged 16 to 21 years (Round 1, n=14; Round 2, n=268) from urban and regional settings in Australia who self-reported that their parent has a mental illness. “Regional” was used to refer to nonurban participants in the study. After ascertaining whether a Web-based intervention was warranted, Web-based intervention features were identified, including how the site might be facilitated, topics, duration and frequency, and the nature of the professional contact. The extent to which young people agreed on the importance of these factors was assessed. Differences and similarities across gender and location were investigated. A mixed method analytic framework was employed using thematic analysis as well as 2-way between-groups analysis of covariance (ANCOVA) controlling for age and chi-square test of independence analysis.

Results: Both rounds highlighted a strong preference for a Web-based intervention. Consensus was reached for a professionally monitored site, young people and professionals having equal input into the weekly facilitated sessions (eg, sharing the lead role in discussions or deciding on relevant session content), unlimited time access, 1-hour, open discussion, weekly sessions over 6 weeks, psychoeducation about mental illness, and considerations for the management of safety violations. There were significant main effects of location type and several of the preferred features for a Web-based intervention for young people who have a parent with a mental illness. However, effect sizes were small to moderate, limiting practical application.

Conclusions: Young people aged 16 to 21 years indicated a need for a professionally monitored, psychoeducational, Web-based intervention, with input from professional facilitators and other young people who have a parent with a mental illness, in addition to recommendations to external resources. These findings may inform the development of future Web-based interventions for this highly vulnerable group.

(J Med Internet Res 2018;20(10):e10158) doi:10.2196/10158

KEYWORDS

Delphi studies; early medical intervention; internet; preventative health; Web-based intervention; youth

Introduction

The transition to adulthood can be a period marked by the onset of serious mental illness. While early intervention can reduce the severity and persistence of these illnesses, treatment is often delayed or not provided, resulting in a potentially avoidable disease burden [1]. Young people who have a parent with a mental illness are a particularly at-risk group, whose risk of developing mental illness ranges from 41% to 77% [2]. Web-based preventative and early interventions hold great promise. However, young people's preferences on how such interventions might meet their needs have not been explored. This study aimed to identify whether these young people want a Web-based intervention and, if so, their perspectives on how it might be delivered and their preferred features.

Compared to their same-aged peers, young people who have a parent with a mental illness face an increased risk of acquiring a substance use disorder or mental illness, academic failure, and developing stress-related somatic health conditions such as asthma [3,4]. These young people may take on caring responsibilities for their parent or siblings [5]. In Australia, 21% to 23% of young people are estimated to have at least 1 parent with a mental illness [6]. Hence, these young people are a prevalent, high-risk group that warrant effective intervention to prevent or reduce the increased risks of adverse outcomes associated with parental mental illness.

Interventions for families impacted by parental mental illness have been shown to be effective. Siegenthaler, Munder, and Egger [7] found that interventions developed for this group decreased young people's risk of developing mental illness by up to 40%. However, most interventions target the parent; for example, *Let's Talk about Children* involves a clinician working with a parent to promote parenting competence and confidence within the context of their illness [8]. Yet, some parents do not seek help because they are in hospital, they do not acknowledge the impact of their illness on children, or are reluctant to disclose their parenting responsibilities because of the fear of losing their children to child protection authorities [8-10]. Some interventions exclude parents who are very ill or persistently use substances [8]. Consequently, the most vulnerable young people may miss out on essential services [11].

There are some youth-specific interventions for parental mental illness that aim to present psychoeducation, provide respite from caring responsibilities, and promote peer connectedness [12]. However, many peer support interventions lack a strong theoretical framework [13], and the evaluation undertaken is generally of relatively poor quality [14]. Many interventions for children in these families have age limits for participation (eg, only for those aged under 18) or exclude young people who have their own mental health issues. Additionally, reliance on public transport, lack of time, parental consent and, especially in regional areas, geographic constraints may impede young people's attendance [15]. Given the stigma associated with mental illness [16], some young people in these families prefer anonymous services such as helplines [17].

To succeed in identifying and supporting young people who have a parent with a mental illness, services need to engage with young people in environments where they seek help and interact [18]. Young people are increasingly turning to the internet to find mental health information and support [19]. In Canada, Wetterlin et al [20] found that 61.6% of 521 young people aged between 17 to 24 years had utilized the internet to access information or seek help for how they were feeling, and 82.9% indicated that they were likely to use a mental health website to find information in challenging times. In a recent review on the use of Web-based and social networking interventions, it was recommended that services promote the use of Web-based technologies given the clear benefits in reducing symptomology (eg, depression) in young people [21]. The internet is appealing to young people because it is anonymous and may be easily accessed at all times of the day [22]. In particular, Web-based interventions may reach a more diverse population than traditional face-to-face interventions [23], including young people who might otherwise avoid services [16]. Grové et al [24] found a strong preference for Web-based support among young people whose parents have a mental illness. Thus, a Web-based medium may be a useful and age-appropriate platform in which to engage and support young people who have a parent with a mental illness.

There are some Web-based interventions specifically designed for this group of young people, including *Survivalkid* [22,25], *Grubbel* [26], and *Kopstoring* [27], although more research is needed on the evaluation and suitability of Web-based platforms for this at-risk target group. When developing mental health interventions, it is important to consult with end users [28]. While some studies have sought views from parents or professionals to ascertain what young people need [29], young people themselves often have differing opinions compared with adults [30]. Hence, it is crucial to solicit young peoples' views on whether they want Web-based interventions and how such a medium may function. Capturing these preferences at the outset of intervention development will ensure that the resulting Web-based intervention is responsive to their needs and likely to be utilized [28].

The aim of this study was to explore the views of young people aged 16 to 21 years who have a parent with a mental illness regarding Web-based interventions. Specific research questions were: (1) Is a Web-based intervention a beneficial way to support this target group? If so, (2) How might a Web-based intervention be facilitated and monitored? (3) What features should be included in a Web-based intervention? (4) What topics (if any) should be included in a Web-based intervention? (5) What safety issues (if any) are there in providing a Web-based approach, and how this might be managed?

As females have been shown to be more interested in Web-based interventions and more likely to seek help compared with males [31], and location may make a difference to how young people interact with Web-based interventions [32], this study examined potential gender and location differences.

Methods

Research Design

The study employed a 2-round Delphi study with a panel of young people who have a parent with a mental illness. The Delphi approach is a structured method to derive consensus on issues for which evidence is scarce. It involves a panel of experts to elicit their direction on the topic. In this study, the panel of experts was young people with lived experience of having a parent with a mental illness [33]. The approach guaranteed participants' anonymity, thereby preventing possible biases. This is important when dealing with sensitive topics such as mental illness [33].

Participants

Young people residing in Australia, aged 16-21 years (inclusive), fluent in English, with a parent or parents with a mental illness (self-reported or diagnosed) were eligible to participate. Participants were from urban or regional settings. "Regional" was used to refer to nonurban participants in the study (ie, inner regional, outer regional, rural, remote, or very remote). Young people aged under 18 years required parental permission to participate. Participants were not required to participate in both rounds.

Recruitment for both rounds of the Web-based questionnaire was conducted via the researchers' professional networks and through organizations designed for young people (eg, help lines or young people's carers groups). The response rate is not reported, as the number of people who received the anonymous link is unknown. All Round 1 participants were invited to participate in Round 2. Participants were paid Aus $20 for their involvement in each round.

Delphi Round 1

In total, 33% (14/43) of the young people who commenced Round 1 were eligible. Reasons for exclusion from Round 1 were incomplete questionnaires (22/43, 51%), failure to meet age criteria (5/43, 12%), and currently living outside of Australia (3/43, 7%). Round 1 participants were aged 16-21 years, mean 19.0 (SD 1.8) years, with no participants aged 17 years. Of them, 64% (9/14) participants were female and 64% (9/14) were from urban locations.

Delphi Round 2

In Round 2, 99.3% (268/270) of the participants, aged 16-21 years, mean 19.4 (SD 1.2) years, were eligible to be included. In total, 28% (12/43) of the participants from Round 1 (ie, who partially or fully completed the questionnaire) participated in Round 2. Other participants for Round 2 (256/268, 95.5%) were invited via professional networks to participate in this study via an anonymous link. Of them, 0.8% (2/268) of the participants were excluded as one was too young and the other was living outside of Australia. Participants aged 20 years (99/268, 36.7%) represented the largest age group. While the largest number of participants were from urban locations (119/268, 44.4%), the representation was more widespread compared to Round 1.

Participants self-reported their parent's main mental illness(es) aligned with the Diagnostic and Statistical Manual of Mental Disorders 5th edition [34]. Parental mental illnesses in Round 1 and Round 2 represented 25% (5/20) and 55% (11/20) of the Diagnostic and Statistical Manual of Mental Disorders 5th edition [34] classification chapters, respectively. Depressive and anxiety disorders were most commonly reported.

Materials and Procedure

The study was approved by the Monash University Human Research Ethics committee. Web-based explanatory statements were provided to participants prior to each round. The questionnaire for Round 1 was based on the research questions and consisted of 22 questions in total, mainly of open-ended responses congruent with the Delphi method [35]. Questions related to different features of a Web-based intervention for young people who have a parent with a mental illness. Round 1 comprised of 14 open-ended questions related to intervention facilitation, interaction with peers and how this may function, preferences regarding communicating with a clinician, possible topics of interest, and safety issues and how such issues may be managed. The Round 1 questionnaire posed a closed question regarding whether a Web-based approach was a useful way to support young people who have a parent with a mental illness, along with an opportunity to provide reasons for a yes or no response.

The Round 2 questionnaire was developed from the themes identified in Round 1. It consisted of 39 questions related to the preferred structure of a Web-based intervention, peer interaction, whether the intervention should be time-limited, and preferred frequency and length of clinician-facilitated sessions. There were another 24 questions that required participants to rate the level of importance of the various features identified in Round 1 on an 11-point Likert-type scale from 0 (not important) to 10 (extremely important). These features related to the benefits of accessing a Web-based intervention, topics for Web-based intervention sessions, how Web-based interventions should be facilitated and monitored, internet safety, and issues around ethical issues (eg, informed consent). As discussed in preliminary data analyses, a consensus was reached after 2 rounds of the Delphi method.

Preliminary Data Analyses

Data were collected using the Qualtrics Web-based questionnaire package and analyzed using IBM SPSS Statistics version 24. Delphi studies typically attract similar, smaller numbers of participants in each round (eg, 12-30 participants). However, the Round 2 participant numbers far exceeded those of Round 1, thus providing an opportunity to examine potentially important differences in participant subgroups according to age, gender, and location.

The Braun and Clarke [36] 6-step analytic approach was employed to analyze Round 1 Delphi data. Themes were identified independently by 2 researchers. The researchers then compared and contrasted their interpretations of the main themes by referring back to participants' responses, to reach a consensus through discussion rather than a numerical level of agreement.

The Round 2 analysis included a rank ordering of preferences, frequencies, and the use of analysis of covariance (ANCOVA) and chi-square test of independence to identify differences across gender and urban and regional locations. In apriori correlations, age was found to be positively correlated with almost all variables, with older participants scoring more highly than those younger. This was subsequently controlled for in the ANCOVAs reported in the results section. Prior to analysis, data were examined to ensure that assumptions pertaining to each of the tests were met, with no violations encountered. Due to the exploratory nature of the analyses, Bonferroni corrections were not adopted (the likelihood of this increasing the type 1 errors is acknowledged) [37], and given the large amount of data generated in Round 2, only the key findings are reported here.

Results

For ease of reporting, Round 1 and Round 2 findings are combined. Descriptive statistics for themes are shown in Table 1, with Round 1 themes in the left-hand column in Tables 2, 3, and 4. The tables are stratified by gender and location. Significant differences, significant interactions between gender and location on themes, and main effects are indicated in the tables and discussed in the text.

All participants in Round 1 (14/14, 100%) agreed that a Web-based intervention was a beneficial way to support young people who have a parent with a mental illness. This was extended in Round 2, where participants rated access to a Web-based intervention as very important (mean 8.52) on the 11-point (0-10) scale. The ANCOVA showed no main or interaction effects for gender or location.

In Round 1, 93% (13/14) participants agreed that a clinician should monitor a Web-based intervention group. In Round 2, participants rated it very important for a clinician to monitor a Web-based intervention (mean 8.01). The ANCOVA revealed neither interaction effect nor main effect for gender; however, there was a significant main effect for location, $F_{1, 240}$=7.44, P=.01, partial η^2=0.030. Urban participants rated a Web-based intervention facilitated by a clinician to be more important (mean 8.21) than regional participants (mean 7.88). Note, however, that both groups provided very high ratings.

Round 2 participants were asked for their preference for the duration (as a percentage) that they wanted a clinician and peers to lead discussions during each Web-based group session. No participant preferred 100% direction from either a clinician or peers. These preferences are shown in Table 1 and were analyzed further (ie, gender and location) using chi-square, with the only significant difference being that more urban young people reported that the approach should include weekly goals to practice between sessions compared with regional participants, χ^2_1=12.4, P<.001, phi=−0.266 (n=244).

As shown in Table 1, Round 2 participants preferred weekly sessions over 6 weeks, an “open and free discussion depending on the needs of the group that week” (ranked 1). Preferences about time limitations, length, frequency, and duration of individual sessions were analyzed further (ie, gender and location) using chi-square, with the only significant difference being that more urban young people reported that the approach should be unlimited in time, compared with regional participants, χ^2_1=5.2, P=.02, phi=0.155 (n=244).

Participants were then asked about the features that should be in a Web-based intervention. As shown in Table 2, Round 1 participants agreed there should be multiple features, and Round 2 participants rated these features as very important.

Table 1. Round 2 first preference rankings and participant frequency analysis for Web-based group intervention structure based on Round 1 themes.

Web-based intervention themes	Round 2 total (N=268)	Gender		Location type	
		Males (n=157)	Females (n=111)	Urban (n=119)	Regional (n=125)
Web-based intervention structure, n (%)					
Unlimited timeframe preferred[a]	201 (75.0)	125 (79.6)	76 (68.5)	97 (81.5)	85 (68.0)
6-week intervention preferred by participants who selected a time-limited intervention	27 (10.1)	14 (8.9)	13 (11.7)	8 (6.7)	19 (15.2)
Weekly session frequency	171 (63.8)	98 (62.4)	73 (65.8)	77 (64.7)	80 (64.0)
Duration of 1 hour	120 (44.8)	67 (42.7)	53 (47.7)	46 (38.7)	61 (48.8)
Group session facilitation, n (%)					
Ratio of 50% by clinician and 50% by peer	138 (51.5)	76 (48.4)	62 (55.9)	67 (56.3)	56 (44.8)
Group session structure (≥1 preferences allowed), n (%)					
Open and free discussion depending on needs of the group that week (ranking 1)	227 (84.7)	126 (80.3)	81 (73.0)	99 (83.2)	17 (13.6)
Set topics each week (ranking 2)	207 (77.2)	134 (85.4)	93 (83.8)	93 (78.2)	92 (73.6)
Weekly goals to work on and practice between sessions[a] (ranking 3)	190 (70.9)	113 (72.0)	77 (69.4)	96 (80.7)	75 (60.0)

[a] Significant difference reported for location.

Table 2. Round 1 themes and Round 2 mean importance ratings and rankings of Round 1 themes for preferred features of a Web-based intervention.

Subthemes identified in Round 1	Round 2 importance rating (N=268)[a]	Location type (n=244)[b]	
		Urban (n=119)	Regional (n=125)
Referral(s) to other sources for additional support[c], mean (SD)	7.84 (1.3)	N/A[d]	N/A
Males	N/A	7.78 (1.4)	7.97 (1.0)
Females	N/A	8.16 (1.3)	7.63 (1.2)
Privately contact clinician from Web-based sessions[e], mean (SD)	7.81 (1.2)	N/A	N/A
Males	N/A	7.72 (1.4)	7.73 (1.1)
Females	N/A	8.08 (1.3)	7.73 (1.1)
Open access to a chat room[c], mean (SD)	7.76 (1.3)	N/A	N/A
Males	N/A	7.52 (1.6)	7.96 (0.9)
Females	N/A	8.06 (1.4)	7.65 (1.3)
Privately contact peers (Web-based intervention group), mean (SD)	7.69 (1.4)	N/A	N/A
Males	N/A	7.64 (1.5)	7.82 (1.2)
Females	N/A	7.76 (1.5)	7.37 (1.4)
Referral(s) to additional clinical support[e], mean (SD)	7.40 (1.2)	N/A	N/A
Males	N/A	7.20 (1.3)	7.35 (1.0)
Females	N/A	7.68 (1.4)	7.49 (1.2)
Web-based chat room sessions with set topics decided by a clinician, mean (SD)	7.31 (1.4)	N/A	N/A
Males	N/A	7.22 (1.5)	7.57 (0.2)
Females	N/A	7.16 (1.8)	7.31 (0.2)
Young people and a clinician in a Web-based chat room session, mean (SD)	7.24 (1.5)	N/A	N/A
Males	N/A	6.99 (1.7)	7.41 (1.4)
Females	N/A	7.54 (1.6)	7.29 (1.2)

[a] 11-point Likert-type scale from 0 (not important) to 10 (extremely important).

[b] Males (143/244, 58.6%); females (101/244, 41.4%).

[c] Significant interaction reported between gender and location.

[d] N/A: not applicable.

[e] Significant interaction effects for gender.

The ANCOVA analysis identified a significant interaction between gender and location on referral(s) to other sources for additional support ($F_{1, 239}$=5.05, P=.03, η^2=0.021) and open access to a chat room ($F_{1, 239}$=5.95, P=.02, η^2=.024), with female urban participants reporting significantly higher scores for these preferred features (Table 2 for scores). Additionally, there were 2 significant interaction effects for gender for privately contacting a clinician from the Web-based sessions ($F_{1, 239}$=4.66, P=.03, η^2=0.019) and referral(s) to additional clinical support ($F_{1, 239}$=7.18, P=.008, η^2=0.029) with urban females scoring more highly for these preferred features (Table 2 for scores).

In Round 1, 79% (11/14) participants indicated that self-care, resilience building, ways of coping, psychoeducation, and how to ask for help and where to get help were important topics to include in a Web-based intervention. They also indicated a range of other topics that, for Round 2, were grouped into 4 common themes and 10 subthemes (Table 3). Table 3 shows the Round 1 topics and descriptive statistics, importance ratings, and rankings from Round 2.

While participants scored all themes and subthemes highly (at 7 and above), they scored highest in regard to helping and supporting a parent with a mental illness, useful ways to cope when parents have a mental illness, and psychoeducation around specific parental mental illness. A series of ANCOVA analyses were then undertaken with all but one (general mental illness) showing main effects for location, with urban participants having higher mean scores than their regional counterparts. For brevity, ANCOVA statistics are not shown here but are available from the authors. In regard to gender, there was only 1 variable (helping and supporting a parent with a mental illness) where males had higher mean scores than females (F-statistics and means not shown here).

Table 3. Round 1 topics of interest with Round 2 descriptive statistics, importance ratings, and rankings.

Themes and subthemes from Round 1 responses and Round 2 gender (males n=143, females n=101)	Importance (N=268)[a]		Location type (n=244)	
	Rating, mean (SD)	Rank	Urban (n=119), mean (SD)	Regional (n=125), mean (SD)
Psychoeducation				
Helping and supporting parent with a mental illness[b,c]	8.69 (1.21)	1	N/A[d]	N/A
Male	N/A	N/A	9.01 (1.0)	8.55 (0.9)
Female	N/A	N/A	8.62 (1.6)	8.31 (1.3)
Specific parental mental illness[b]	8.59 (1.14)	3	N/A	N/A
Male	N/A	N/A	8.86 (1.0)	8.41 (1.0)
Female	N/A	N/A	8.64 (1.4)	8.29 (2.0)
Personal mental health and related information[b]	8.55 (1.20)	4	N/A	N/A
Male	N/A	N/A	8.72 (1.2)	8.34 (1.1)
Female	N/A	N/A	8.76 (1.4)	8.31 (1.1)
Dealing with situations arising from parental mental illness[b]	7.92 (1.26)	8	N/A	N/A
Male	N/A	N/A	8.20 (1.2)	7.69 (1.9)
Female	N/A	N/A	8.26 (1.5)	7.61 (2.0)
General mental illness	7.54 (1.86)	10	N/A	N/A
Male	N/A	N/A	7.74 (1.9)	7.42 (1.7)
Female	N/A	N/A	7.66 (2.1)	7.08 (1.8)
Emotional well-being				
Build resilience[b]	8.54 (1.2)	5	N/A	N/A
Male	N/A	N/A	8.72 (1.2)	8.41 (1.2)
Female	N/A	N/A	8.70 (1.1)	8.14 (1.2)
Promote general emotional well-being[b]	8.45 (1.4)	6	N/A	N/A
Male	N/A	N/A	8.61 (1.6)	8.16 (1.3)
Female	N/A	N/A	8.78 (1.2)	8.12 (1.3)
Self-care strategies[b]	8.38 (1.3)	7	N/A	N/A
Male	N/A	N/A	8.57 (1.3)	8.11 (1.1)
Female	N/A	N/A	8.80 (1.2)	7.96 (1.5)
Coping strategies and skills				
Useful ways to cope when parents have a mental illness[b]	8.65 (1.3)	2	N/A	N/A
Male	N/A	N/A	8.90 (1.1)	8.43 (1.2)
Female	N/A	N/A	8.82 (1.4)	8.31 (1.3)
Other resources				
Access to other resources besides the Web-based intervention[b]	7.58 (1.1)	9	N/A	N/A
Male	N/A	N/A	7.87 (1.1)	7.38 (1.0)
Female	N/A	N/A	7.70 (1.5)	7.31 (.8)

[a] 11-point Likert-type scale from 0 (not important) to 10 (extremely important).

[b] Main effect for location.

[c] Main effect for gender.

[d] N/A: not applicable.

Table 4. Round 1 themes and subthemes with Round 2 importance ratings and rankings for safety issues with a Web-based intervention.

Themes and subthemes from Round 1 responses and gender (males n=143, females n=101)	Round 2 importance (N=268)[a]		Location type (n=244)	
	Mean (SD)	Rank	Urban (n=119), mean (SD)	Regional (n=125), mean (SD)
Knowledge of prohibition from Web-based intervention if safety guidelines are not adhered to	7.60 (1.34)	1	N/A[b]	N/A
Males	N/A	N/A	7.42 (1.53)	7.78 (1.13)
Females	N/A	N/A	7.70 (1.39)	7.69 (1.26)
Clinician monitoring for cyberbullying	7.59 (1.47)	2	N/A	N/A
Males	N/A	N/A	7.51 (1.61)	7.66 (1.19)
Females	N/A	N/A	7.84 (1.46)	7.63 (1.26)
Web-based communication	7.53 (1.38)	3	N/A	N/A
Males	N/A	N/A	7.41 (1.43)	7.55 (1.15)
Females	N/A	N/A	7.80 (1.47)	7.57 (1.53)
Participation without parental consent (n=56)	6.48 (2.14)	N/A	N/A	N/A
Males (n=26)	N/A	N/A	6.57 (1.3)	6.37 (1.2)
Females (n=30)	N/A	N/A	6.80 (1.2)	6.70 (1.1)

[a] 11-point Likert-type scale from 0 (not important) to 10 (extremely important).

[b] N/A: not applicable.

Participants raised 2 potential safety and ethical issues, specifically in regard to informed consent. Participants did not want to have to obtain parental consent for several reasons: a perception that their parent would not understand why they sought help, a concern about embarrassing or disappointing their parent, fear of their parent's reaction (eg, anger, guilt, or upset), worry that their parents would ban it, and a concern that by asking, they would strain the relationship with their parent, particularly if the parent with the mental illness was the only parent in the house. The key themes from Round 1 responses importance ratings and rankings for safety issues with a Web-based intervention are shown in Table 4. An ANCOVA revealed no interaction effect or main effects for gender or location.

Discussion

Principal Findings

The young people involved in this Delphi study (across age groups and locations) indicated a strong preference for a Web-based approach along with a range of preferred features. Developers should attempt to take these preferences into account when developing Web-based interventions for this target group. An especially important finding was young people's preferences for a Web-based intervention that was facilitated by professionals. This preference has efficacy support; in a meta-analysis, Andersson and Cuijpers [38] found that Web-based interventions with professional support were much more effective than those without professional support in the treatment of adult depression.

Notwithstanding their preference for professional facilitation, young people preferred to have equal input during weekly sessions, with young people and professional facilitators sharing the role of directing discussions and contributing to session content. For example, a professional facilitator may allow young people to spontaneously direct discussions around their preferred content for up to half of the weekly session based on their needs or interest. Young people also indicated a preference for weekly, scheduled, 1-hour open discussion sessions, according to the particular needs of the group, as well as set topics. Hence, it appears to be important that Web-based interventions are sufficiently flexible to present established topics and be responsive to the immediate needs of young people participating in the intervention. The *Kopstoring* Web-based intervention has pre-established topics each week but does not appear to allow for open-ended discussion sessions [27]. Additionally, young people indicated a preference for weekly goals to work on between sessions. Others have also highlighted the importance of goal setting with these young people; one study found that through goal setting, young people who have a parent with a mental illness were more likely to achieve their personal goals, which included increasing their understanding of mental illness and improving family connections [39].

Psychoeducation was considered the most important topic to be covered by the intervention. In particular, young people identified a need to learn about their parent's mental illness, how to support their parent, personal mental health, and how they might deal with situations arising from their parent's illness. These findings respond to reports of young people's limited understanding of their parent's mental illness [40]. Psychoeducation is a key component of a range of interventions for families affected by parental mental illness [3,41] and resonates with previous findings regarding children's requests to find out more about their parent's illness [24,42]. Young people also wanted to learn adaptive coping strategies, which is important given the tendency for some young people in these families to employ coping strategies such as avoidance or isolation [43].

As well, young people indicated a preference for a time-unlimited intervention. If it had to be time-limited, young people preferred a 6-week Web-based intervention. Some Web-based interventions for this target group are offered weekly and are generally limited to 8 weeks [26,27]. Survivalkid, in the Netherlands, is unlimited but does not offer structured weekly topics [22]. Further research is required to ascertain the most effective length of Web-based interventions for this target group in terms of efficacy as well as participant satisfaction.

In Round 2, young people identified a need for referrals for additional support. Likewise, others have suggested that multiple interventions may need to be made available to this group of young people [44]. Empowering young people to access their own supports (rather than rely on their parents) potentially addresses the service gaps with identification, referrals, and fears around accessing help due to stigma [45]. Urban females scored higher than all others on referrals for support and wanting open access to chat rooms. This is not surprising given young female adults have reported themselves as their strongest help-seeking influence compared with other social influences, and this was higher for females compared with males [31]. Additionally, young people wanted the opportunity to privately chat with a Web-based facilitator. Survivalkid in the Netherlands has "survival coaches," not therapists, who provide advice and may refer young people to more formal services if required [25]. Organizations would need to consider the viability of private contact given the time demands, staffing required, and costs involved.

Lastly, young people were asked to identify potential safety issues around providing a Web-based approach and invited to consider how this could then be managed. Young people wanted assurances that participants not adhering to guidelines would be banned or otherwise dealt with appropriately. Likewise, internet safety is an important consideration for young people when engaging in Web-based interventions [20,46]. Young people aged 16 and 17 indicated that they wanted the opportunity to participate in a Web-based intervention without having to obtain their parent's consent, primarily because they were concerned about their parent's reaction. Parents may act as "gatekeepers" for their children's involvement in interventions [47], a concern aligned with the views of young people elicited here. Finally, it is a normal developmental phase for young adults to differentiate themselves from their parents [48], which again is well encapsulated in the preferences of young people in this study.

Generally, there were statistically significant main effects of location type between urban and regional young people. However, the practical significance [37] of this is limited given the small effect sizes and the observation that actual importance ratings for the different variables were still high for both urban and regional young people.

Limitations

Several limitations apply in interpreting the findings. A smaller number of young people aged 16 and 17 years were recruited compared with the other age groups. Parental mental illnesses were self-reported by those surveyed rather than clinically diagnosed. It is unknown whether participants had their own mental health concerns, which may impact participants' preferences for a Web-based intervention. The Web-based recruitment strategy might have favored those young people with an existing interest in and preference for Web-based supports. Future studies might differentiate between specific age groups to identify their respective needs for Web-based support, who might also be screened for their own mental health concerns. Consideration is also required to develop strategies for identifying young people at an earlier age. Working in collaboration with parents will be key here. Moreover, ongoing research is needed to investigate and evaluate the efficacy and effectiveness of Web-based interventions. The uptake of Web-based interventions also needs to be gauged, given that others have found that while young people prefer Web-based interventions, engagement is often low [20].

Conclusions

This study constitutes a step in the development of Web-based interventions for young people who have a parent with a mental illness. Young people indicated a preference for a time-unlimited, Web-based intervention that was professionally monitored but with input from both young people and professionals. The results of this study might be used to inform the development of Web-based interventions in regard to topics, intervention length and moderation, and safety requirements. The value in seeking the views of those with lived experience is vital in the design of support. Web-based interventions hold promise to prevent or reduce the risk of adverse outcomes in young people into adulthood and potentially reduce the intergenerational risks of parental mental illness on future generations.

Conflicts of Interest

None declared.

References

1. McGorry PD, Hamilton MP. Stepwise expansion of evidence-based care is needed for mental health reform. Med J Aust 2016 May 16;204(9):351-353. [Medline: 27169969]

2. Hosman CMH, van Doesum KTM, van Santvoort F. Prevention of emotional problems and psychiatric risks in children of parents with a mental illness in the Netherlands: I. The scientific basis to a comprehensive approach. Australian e-Journal for the Advancement of Mental Health 2014 Dec 17;8(3):250-263. [doi: 10.5172/jamh.8.3.250]
3. Farahati F, Marcotte D, Wilcox-Gök V. The effects of parents' psychiatric disorders on children's high school dropout. Economics of Education Review 2003 Apr;22(2):167-178. [doi: 10.1016/S0272-7757(02)00031-6]
4. Reupert A, Maybery D. What do we know about families where parents have a mental illness? A systematic review. Child & Youth Services 2016 Apr 14;37(2):98-111. [doi: 10.1080/0145935X.2016.1104037]
5. Kallander EK, Weimand BM, Becker S, Van Roy B, Hanssen-Bauer K, Stavnes K, et al. Children with ill parents: extent and nature of caring activities. Scand J Caring Sci 2017 Sep 04:1-12. [doi: 10.1111/scs.12510] [Medline: 28869667]
6. Maybery DJ, Reupert AE, Patrick K, Goodyear M, Crase L. Prevalence of parental mental illness in Australian families. Psychiatr. bull 2018 Jan 2;33(01):22-26. [doi: 10.1192/pb.bp.107.018861]
7. Siegenthaler E, Munder T, Egger M. Effect of preventive interventions in mentally ill parents on the mental health of the offspring: systematic review and meta-analysis. J Am Acad Child Adolesc Psychiatry 2012 Jan;51(1):8-17.e8. [doi: 10.1016/j.jaac.2011.10.018] [Medline: 22176935]
8. Solantaus T, Paavonen EJ, Toikka S, Punamäki R. Preventive interventions in families with parental depression: children's psychosocial symptoms and prosocial behaviour. Eur Child Adolesc Psychiatry 2010 Dec;19(12):883-892 [FREE Full text] [doi: 10.1007/s00787-010-0135-3] [Medline: 20890622]
9. Foster KP, Hills D, Foster KN. Addressing the support needs of families during the acute hospitalization of a parent with mental illness: A narrative literature review. Int J Ment Health Nurs 2018 Apr;27(2):470-482. [doi: 10.1111/inm.12385] [Medline: 28929572]
10. David DH, Styron T, Davidson L. Supported Parenting to Meet the Needs and Concerns of Mothers with Severe Mental Illness. Am J Psychiatr Rehabil 2011 Apr;14(2):137-153 [FREE Full text] [doi: 10.1080/15487768.2011.569668] [Medline: 22180730]
11. Potter R, Mars B, Eyre O, Legge S, Ford T, Sellers R, et al. Missed opportunities: mental disorder in children of parents with depression. Br J Gen Pract 2012 Jul;62(600):e487-e493 [FREE Full text] [doi: 10.3399/bjgp12X652355] [Medline: 22781997]
12. Foster K, McPhee I, Fethney J, McCloughen A. Outcomes of the ON FIRE peer support programme for children and adolescents in families with mental health problems. Child & Family Social Work 2014 Mar 20;21(3):295-306. [doi: 10.1111/cfs.12143]
13. Reupert AE, Cuff R, Drost L, Foster K, van Doesum KTM, van Santvoort F. Intervention programs for children whose parents have a mental illness: a review. Med J Aust 2013 Aug 05;199(3 Suppl):S18-S22. [Medline: 25369843]
14. Bee P, Bower P, Byford S, Churchill R, Calam R, Stallard P, et al. The clinical effectiveness, cost-effectiveness and acceptability of community-based interventions aimed at improving or maintaining quality of life in children of parents with serious mental illness: a systematic review. Health Technol Assess 2014 Feb;18(8):1-250 [FREE Full text] [doi: 10.3310/hta18080] [Medline: 24502767]
15. Van Doesum KTM, Riebschleger J, Carroll J, Grové C, Lauritzen C, Mordoch E, et al. Successful recruitment strategies for prevention programs targeting children of parents with mental health challenges: An international study. Child & Youth Services 2016 Mar 30;37(2):156-174. [doi: 10.1080/0145935X.2016.1104075]
16. Oh E, Jorm AF, Wright A. Perceived helpfulness of websites for mental health information: a national survey of young Australians. Soc Psychiatry Psychiatr Epidemiol 2009 Apr;44(4):293-299. [doi: 10.1007/s00127-008-0443-9] [Medline: 18830552]
17. Grové C, Reupert A, Maybery D. Peer connections as an intervention with children of families where a parent has a mental illness: Moving towards an understanding of the processes of change. Children and Youth Services Review 2015 Jan;48:177-185. [doi: 10.1016/j.childyouth.2014.12.014]
18. Marcus MA, Westra HA, Eastwood JD, Barnes KL, Mobilizing Minds Research Group. What are young adults saying about mental health? An analysis of Internet blogs. J Med Internet Res 2012 Jan 30;14(1):e17 [FREE Full text] [doi: 10.2196/jmir.1868] [Medline: 22569642]
19. Gowen LK. Online Mental Health Information Seeking in Young Adults with Mental Health Challenges. Journal of Technology in Human Services 2013 Apr;31(2):97-111. [doi: 10.1080/15228835.2013.765533]
20. 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]
21. Rice SM, Goodall J, Hetrick SE, Parker AG, Gilbertson T, Amminger GP, et al. Online and social networking interventions for the treatment of depression in young people: a systematic review. J Med Internet Res 2014;16(9):e206 [FREE Full text] [doi: 10.2196/jmir.3304] [Medline: 25226790]
22. Drost LM, Cuijpers P, Schippers GM. Developing an interactive website for adolescents with a mentally ill family member. Clin Child Psychol Psychiatry 2011 Jul;16(3):351-364. [doi: 10.1177/1359104510366281] [Medline: 20980365]
23. Griffiths KM, Christensen H. Internet-based mental health programs: a powerful tool in the rural medical kit. Aust J Rural Health 2007 Apr;15(2):81-87. [doi: 10.1111/j.1440-1584.2007.00859.x] [Medline: 17441815]

24. Grové C, Reupert A, Maybery D. The Perspectives of Young People of Parents with a Mental Illness Regarding Preferred Interventions and Supports. J Child Fam Stud 2016 Jun 18;25(10):3056-3065. [doi: 10.1007/s10826-016-0468-8]
25. Drost LM, Schippers GM. Online support for children of parents suffering from mental illness: a case study. Clin Child Psychol Psychiatry 2015 Jan;20(1):53-67. [doi: 10.1177/1359104513496260] [Medline: 23904177]
26. Elgán TH, Kartengren N, Strandberg AK, Ingemarson M, Hansson H, Zetterlind U, et al. A web-based group course intervention for 15-25-year-olds whose parents have substance use problems or mental illness: study protocol for a randomized controlled trial. BMC Public Health 2016 Sep 23;16(1):1011 [FREE Full text] [doi: 10.1186/s12889-016-3691-8] [Medline: 27663995]
27. Woolderink M, Smit F, van der Zanden R, Beecham J, Knapp M, Paulus A, et al. Design of an internet-based health economic evaluation of a preventive group-intervention for children of parents with mental illness or substance use disorders. BMC Public Health 2010 Aug 10;10:470 [FREE Full text] [doi: 10.1186/1471-2458-10-470] [Medline: 20698962]
28. Trowse LC, Cook JG, Clooney TJA. Mental health consumer and carer participation: why we bother. MJA Open 2012 Apr 16;1(Suppl 1):10. [doi: 10.5694/mjao11.11141]
29. Jander A, Crutzen R, Mercken L, De Vries H. Web-based interventions to decrease alcohol use in adolescents: a Delphi study about increasing effectiveness and reducing drop-out. BMC Public Health 2015 Apr 09;15:340 [FREE Full text] [doi: 10.1186/s12889-015-1639-z] [Medline: 25881254]
30. Maybery D, Ling L, Szakacs E, Reupert A. Children of a parent with a mental illness: perspectives on need. Australian e-Journal for the Advancement of Mental Health 2014 Dec 17;4(2):78-88. [doi: 10.5172/jamh.4.2.78]
31. Rickwood DJ, Mazzer KR, Telford NR. Social influences on seeking help from mental health services, in-person and online, during adolescence and young adulthood. BMC Psychiatry 2015;15:40 [FREE Full text] [doi: 10.1186/s12888-015-0429-6] [Medline: 25886609]
32. Griffiths KM, Mackinnon AJ, Crisp DA, Christensen H, Bennett K, Farrer L. The effectiveness of an online support group for members of the community with depression: a randomised controlled trial. PLoS One 2012 Dec;7(12):e53244 [FREE Full text] [doi: 10.1371/journal.pone.0053244] [Medline: 23285271]
33. Hsu C, Sandford SB. The Delphi technique: Making sense of consensus. Practical Assessment, Research & Evaluation 2007;12(10):1-8.
34. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. Washington, DC: Author; 2013.
35. Custer RL, Scarcella JA, Stewart BR. The modified Delphi technique: A rotational modification. Journal of Career and Technical Education 1999;15(2):50-58.
36. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
37. Gravetter FG, Wallnau LB. Statistics for the Behavioural Sciences (8th ed). In: Statistics for the Behavioural Sciences. California: Wadsworth; 2009.
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. Maybery D, Reupert A, Goodyear M. Goal setting in recovery: families where a parent has a mental illness or a dual diagnosis. Child & Family Social Work 2013 Jul 23;20(3):354-363. [doi: 10.1111/cfs.12084]
40. Mordoch E. How children understand parental mental illness: “you don't get life insurance. What's life insurance?”. J Can Acad Child Adolesc Psychiatry 2010 Feb;19(1):19-25 [FREE Full text] [Medline: 20119563]
41. Marston N, Stavnes K, Van Loon LMA, Drost LM, Maybery D, Mosek A, et al. A content analysis of Intervention Key Elements and Assessments (IKEA): What's in the black box in the interventions directed to families where a parent has a mental illness? Child & Youth Services 2016 Feb 24;37(2):112-128. [doi: 10.1080/0145935X.2016.1104041]
42. Gladstone BM, Boydell KM, Seeman MV, McKeever PD. Children's experiences of parental mental illness: a literature review. Early Interv Psychiatry 2011 Nov;5(4):271-289. [doi: 10.1111/j.1751-7893.2011.00287.x] [Medline: 21883973]
43. Holmila MJ, Itäpuisto M, Ilva M. Invisible victims or competent agents: Opinions and ways of coping among children aged 12-18 years with problem drinking parents. Drugs: Education, Prevention and Policy 2010 Oct 05;18(3):179-186. [doi: 10.3109/09687637.2010.493168]
44. Steer S, Reupert A, Maybery D. Programs for Children of Parents who have a Mental Illness: Referral and Assessment Practices. “One size fits all”? Australian Social Work 2011 Dec;64(4):502-514. [doi: 10.1080/0312407X.2011.594901]
45. Trondsen MV. Living with a mentally ill parent: exploring adolescents' experiences and perspectives. Qual Health Res 2012 Feb;22(2):174-188. [doi: 10.1177/1049732311420736] [Medline: 21873281]
46. Mar MY, Neilson EK, Torchalla I, Werker GR, Laing A, Krausz M. Exploring e-Mental Health Preferences of Generation Y. Journal of Technology in Human Services 2014 Nov 26;32(4):312-327. [doi: 10.1080/15228835.2014.943457]
47. Reupert A, Goodyear M, Maybery D. Engaging with, and understanding children whose parents have a dual diagnosis. Child Adolesc Ment Health 2011 Sep 19;17(3):153-160. [doi: 10.1111/j.1475-3588.2011.00622.x]
48. Kessler RC, Angermeyer M, Anthony JC, Demyttenaere K, Gasquet I, Gluzman S, et al. Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. World Psychiatry 2007 Oct;6(3):168-176 [FREE Full text] [Medline: 18188442]

Abbreviations

ANCOVA: analysis of covariance

Edited by G Eysenbach; submitted 16.02.18; peer-reviewed by D Rickwood, K Machin; comments to author 29.03.18; revised version received 16.05.18; accepted 21.08.18; published 31.10.18.

Please cite as:
Matar JL, Maybery DJ, McLean LA, Reupert A
Web-Based Health Intervention for Young People Who Have a Parent with a Mental Illness: Delphi Study Among Potential Future Users
J Med Internet Res 2018;20(10):e10158
URL: http://www.jmir.org/2018/10/e10158/
doi:10.2196/10158
PMID:30381281

Original Paper

Drug Repositioning to Accelerate Drug Development Using Social Media Data: Computational Study on Parkinson Disease

Mengnan Zhao[1], MSc; Christopher C Yang[1], PhD

[1] College of Computing and Informatics, Drexel University, Philadelphia, PA, United States

Corresponding Author:
Christopher C Yang, PhD
College of Computing and Informatics
Drexel University
3401 Market Street
Philadelphia, PA,
United States
Phone: 1 267 648 4077
Email: ccy24@drexel.edu

Abstract

Background: Due to the high cost and low success rate in new drug development, systematic drug repositioning methods are exploited to find new indications for existing drugs.

Objective: We sought to propose a new computational drug repositioning method to identify repositioning drugs for Parkinson disease (PD).

Methods: We developed a novel heterogeneous network mining repositioning method that constructed a 3-layer network of disease, drug, and adverse drug reaction and involved user-generated data from online health communities to identify potential candidate drugs for PD.

Results: We identified 44 non-Parkinson drugs by using the proposed approach, with data collected from both pharmaceutical databases and online health communities. Based on the further literature analysis, we found literature evidence for 28 drugs.

Conclusions: In summary, the proposed heterogeneous network mining repositioning approach is promising for identifying repositioning candidates for PD. It shows that adverse drug reactions are potential intermediaries to reveal relationships between disease and drug.

(J Med Internet Res 2018;20(10):e271) doi:10.2196/jmir.9646

KEYWORDS

drug repositioning; Parkinson disease; heterogeneous network; social media

Introduction

Parkinson disease (PD) is a chronic and progressive movement disorder with the main symptoms including tremor, slowed movement, rigidity, and postural instability. It is the second most common neurodegenerative disorder and affects more than 5 million people around the world. So far, the causes of PD remain unknown but are thought to be triggered by genetic or environmental factors that lead to the death of dopamine generating cells [1]. Current treatments and medications cannot reverse the effects of the disease but can manage the symptoms; however, they come with many serious adverse drug events. Therefore, it is important to develop new drugs for PD.

De novo drug development has been a prolonged process with very high development costs and a low success rate in the past decades. Meanwhile, drug repositioning, which identifies new indications for existing drugs, is suggested to be a more cost- and time-efficient strategy by the evidence. One main advantage is that since repositioning drugs have already been validated by pharmaceutical and toxicological tests, the time and cost for development and the risk of failure in early stages can be reduced markedly. Currently, systematic drug repositioning methods can be categorized as either disease-based, where discovery initiates from the clinical perspective of diseases, or drug-based, where discovery initiates from the chemical or pharmaceutical perspective of drugs [2]. Drug-based strategies are preferred when there is expertise or capability in using and understanding pharmacological properties of drugs or if rich pharmacological or chemical data for drugs are available. Disease-based strategies are preferred to overcome missing knowledge in the pharmacology of a drug or when repositioning efforts are to be focused on a specific disease or therapeutic category [3]. In this context, we proposed a disease-based approach to identify the repositioning drugs for diseases using social media data and pharmaceutical databases. Specifically, we connected drugs with their potential indications via adverse drug reactions (ADRs).

The phenome is emerging as a new source of information for drug repositioning because phenotypic information is shown to have great potential for detecting novel associations between drugs and diseases. Clinical side effects, which are capable of profiling drug-related phenotypic information, have been found to be helpful in discovering new therapeutic uses for drugs [4]. As a result, the ADR is becoming an important intermediary to connect drugs with diseases in drug repositioning because it reflects the physiological consequences and phenotypic expressions of the drugs. The rationale behind the ADR-based approaches of drug repositioning is based on the underlying mechanism of action linking ADRs with diseases when the ADR is shared by a number of drugs indicated for the disease [5].

Considering the huge potential of social media data for health care informatics, we collected user-generated content from an online health community as one data source. It has been proven that social media data provide a large volume of timely information contributed by health consumers, and the detection of adverse drug reactions using social media can achieve better results than detection using other traditional data sources such as spontaneous reporting systems or electronic health records [6-9]. As network-based approaches have been increasingly used to identify repositioning drug candidates in both drug-based and disease-based strategies [10-12] and heterogeneous networks can capture more essential and accurate features of the health care information compared with a homogeneous network, we exploited the heterogeneous networking mining methodology to reveal relationships between different medical entities.

Methods

Our method mainly comprises the following steps: (1) extracting ADR entities from user posts in online health communities; (2) computing association strength of ADR-drug, ADR-disease, and disease-drug pairs; (3) building heterogeneous networks based on association strength; and (4) identifying significant ADR-disease paths from the network and conducting drug repositioning. Figure 1 shows the workflow of our method.

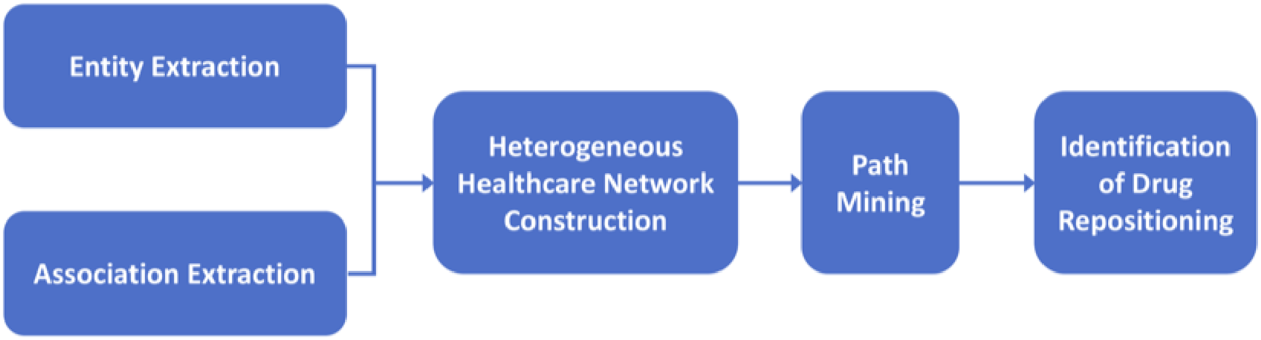
Adverse Drug Reaction Entity Extraction

A significant challenge existing in extracting ADR terms from social media data is the diverse expressions of adverse reactions. Health consumers describe their health issues using different vocabularies than medical professionals due to their lack of medical training. As a result, standard medical lexicons used by professionals like the US National Library of Medicine's Unified Medical Language System are not applicable in analyzing health consumer–contributed content. To address this problem, we used the Consumer Health Vocabulary (CHV) Wiki [13] to build our ADR lexicon. The CHV Wiki links everyday health-related words to professional terms and jargon, and the goal is to bridge the communication gap between consumers and health care professionals [5]. It provides a list of preferred names of ADRs and the corresponding consumer-contributed expressions to each of them. For example, "anorexia" is a professional expression of an ADR, and the CHV Wiki extends it to "appetite lost," "appetite loss," "appetite lack," "no appetite," and several other common expressions. In our study, we used all of the expressions suggested by the CHV Wiki to extract ADR entities in user-generated information [14].

Association Strength Calculation

After extracting ADR entities, we computed the association strength between ADR-disease, ADR-drug, and drug-disease pairs. In social media data, if 2 entities are mentioned together frequently, they are considered to be strongly associated [15]. We conducted co-occurrence analysis on those entity pairs to evaluate their association strength. In co-occurrence analysis, choosing proper analysis granularity is important. That means the co-occurrence analysis should be calculated within an appropriate scale. A post or comment might be set as an analysis unit, but they are usually very short and the users may jump into their problems without describing the context. Thus, a post or comment is too small an analysis unit to calculate the co-occurrence between entities. Instead, a thread, which contains a post and all the following comments, is more appropriate to be an analysis unit because it embodies all the discussions on a particular issue and contains all the keywords. Therefore, we calculated the co-occurrence frequency of entity pairs within each thread.

Figure 1. Workflow of drug repositioning for Parkinson disease based on adverse drug reactions.



We applied association rule mining to determine the co-occurrence frequency between ADR-PD, ADR-drug, and drug-PD pairs. Mathematically, let $I=\{I_1, I_2,..., I_m\}$ be a set of items and let $T=\{T_1, T_2,…, T_n\}$ be a set of transactions, where each transaction is a subset of items such that $T_i \subseteq I$. The association rule is an implication of the form $A \Rightarrow B$, where $A \subset I$, $B \subset I$, and $A \cap B=\emptyset$, which is referred to as an itemset. An itemset that contains *k* items is a *k*-itemset. Specifically, there are both 1-itemsets ({ADR}, {drug(R)}, {PD}) and 2-itemsets ({ADR, R}, {ADR, PD}, {R, PD}) in our experiments.

Lift is a commonly used measure based on probability theory in association rule mining. For instance, *lift* measures the strength of an association by considering not only the co-occurrence of $R \cup ADR$ but also the correlation between itemsets {*R* } and {*ADR* }. To be specific, it computes the ratio of the proportion of threads containing both *R* and *ADR* above the expectation that *R* and *ADR* are independent of each other. The following shows the equations for computing *Lift* ($R \Rightarrow ADR$):

In which *count* (*ADR*) is the number of threads that contain target *ADR*, *count* ($R \cup ADR$) is the number of threads that contain both drug *R* and *ADR*, and *total count* is the total number of threads. Likewise, we can compute *Lift* ($D \Rightarrow ADR$) and *Lift* ($D \Rightarrow R$) with the same equation. The *lift* value is consistent with the strength of associations between 2 items.

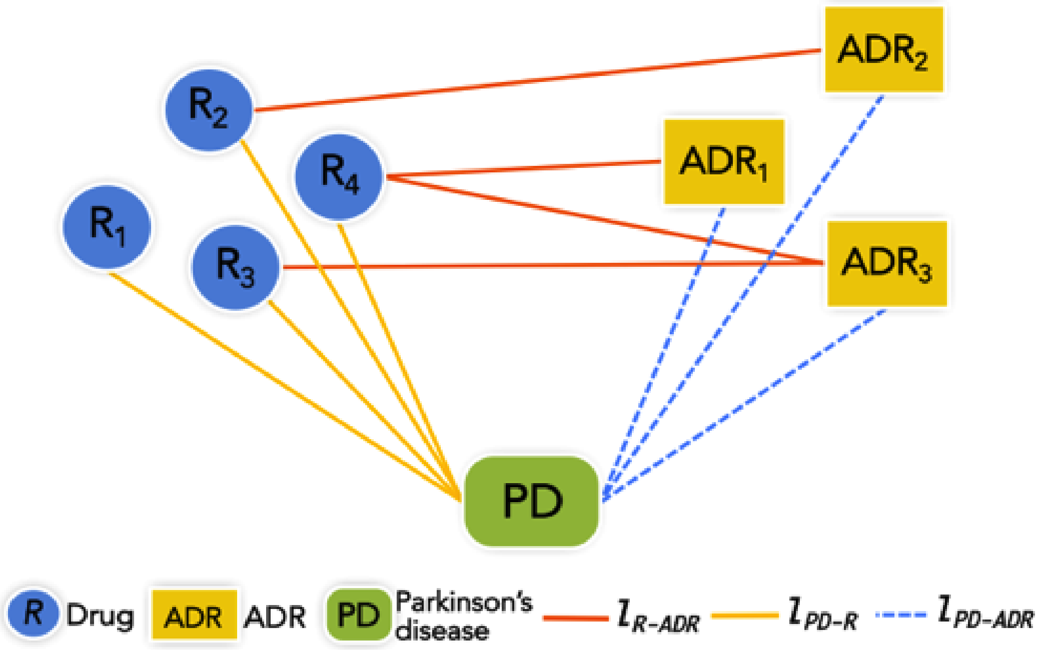
Heterogeneous Network Construction

A heterogeneous network is defined as a graph consisting of nodes connected by links, with at least 2 types of nodes and at least 2 types of links [16]. Let $N=\{n_1,n_2,...,n_k\}$ be a set of nodes and $L=\{l_1,l_2,...,l_m\}$ be a set of links, and $G=(N, L)$ denotes the graph. In the graph *G*, each node $n_i \in N$ belongs to a particular type from γ and each link $l_i \in L$ belongs to a particular type from τ, where $|\gamma|>1$ or $|\tau|>1$.

In our heterogeneous health care network for PD, there are 3 types of nodes (drug, PD, and ADR) and 3 types of links (drug-ADR, PD-ADR, and drug-PD), that is, $\gamma=\{R, PD, ADR\}$ and $\tau=\{L_{\text{R-ADR}}, L_{\text{PD-ADR}}, L_{\text{PD-R}}\}$. Figure 2 presents a nondirectional graph model of the described heterogeneous health care network that is weighted, and the association of the 2 end nodes of a link determines the weight of the link. For instance, the weight of the link between a drug and an ADR is represented and computed by *Lift* ($R \Rightarrow ADR$).

In addition, we predicted the strength of association between disease-ADR pairs through all possible paths suggested by the network in this work and proposed a path mining method: *Path* ($D\text{-}R\text{-}ADR, W_{\text{D-R}} * W_{\text{R-ADR}}$). ADR represents the harmful and unpleasant reactions of medicine use, and hence, the ADRs that are associated with a disease are highly influenced by the drugs that are used for the disease. We use drug as the bridge between disease and ADR. The weight of a path is computed by the products of *Lift* ($D \Rightarrow R$) and *Lift* ($R \Rightarrow ADR$).

Figure 2. Heterogeneous health care network model.



The association between disease and ADR is then computed by the equation seen in below, in which *P* denotes all of the possible paths between *D* and *R*:

Drug Repositioning Based on Disease–Adverse Drug Reaction Association

The basic hypothesis of drug repositioning is that if ADR *X* is significantly associated with disease *D*, then drugs that list *X* as a side effect should be evaluated as candidates for treating disease *D* [5]. For all the disease-ADR associations we detected, we first used a sample *t* test to determine whether they were significant and then applied drug repositioning to those significant associations.

The repositioning process is completed in 3 steps: (1) for each ADR contained in those significant ADR-disease associations, we consulted the Side Effect Resource (SIDER) database [17] to identify the drugs that list the identified ADR as a frequent side effect, (2) we directed those drugs to diseases via the drug-ADR-disease paths and obtained the final drug-disease associations, and (3) we removed the drug-disease pairs that have already been published in the Pharmacogenomics Knowledgebase (PharmGKB) [18]. Thus, the remaining drug-disease pairs could be considered as repositioning drug–disease associations and reported as the drug repositioning results.

Results

Data Collection

In our experiment, user-generated data were collected from MedHelp [19], one of the pioneers in online health communities. Since its introduction in 1994, MedHelp has been the exemplar in online health communities. Currently, it empowers over 12 million people each month to describe and discuss their health and medical problems and find answers. We collected data from MedHelp by implementing an automatic Web crawler that went through MedHelp webpages in HTML format page by page and retrieved information including user name, post, comment, and timestamp. The returned data were saved in .txt files and organized thread by thread, where each thread contained the original post and all the following comments. We collected 12,571 threads (containing 504,097 comments) by using Parkinson and the corresponding 55 drugs suggested by PharmGKB as query words on October 10, 2016. After extracting disease, drug, and ADR entities from the consumer-contributed content, we constructed a heterogeneous network containing 3 types of nodes and 3 types of links. In total, the network contained 255 nodes: 55 drugs, 199 ADRs, and 1 disease.

Repositioning Drugs for Parkinson Disease

When the drugs treating a disease share a common ADR, the ADR and the disease are considered to be connected via an underlying mechanism of action. By applying the proposed heterogeneous network mining method, we detected the ADRs associated with PD and filtered the significant associations based on statistical analysis. As a result, we found 9 ADRs significantly associated with PD: muscle cramp, gastrointestinal disorder, nervous system disorder, angiopathy, somnolence, orthostatic hypotension, carpal tunnel syndrome, hallucination, and influenza.

We retrieved the drugs listing the ADRs but not indicated for PD from the SIDER database with several filtering conditions: the drugs indicate 1 of the 9 ADRs as “frequent,” “common,” or “postmarketing” or with an occurring percentage higher than 10. As a result, we identified 44 repositioning drugs for PD. Table 1 lists all the repositioning drugs, as well as whether these drugs are associated with PD based on literature analysis. By using drug names and Parkinson as queries in PubMed, we then investigated all the findings to understand the implications behind the associations of disease and repositioning drugs. We found evidence in PubMed literature to support a positive effect of some drugs on PD, evidence to support negative or adverse effects of some drugs on PD, evidence that some drugs exert both positive and negative effects on PD, and no explicit association between some drugs and PD.

Multimedia Appendix 1 presents the repositioning drugs that are supported by evidence from PubMed literature, and each row illustrates the drug and how it is labeled with the ADRs in the SIDER database. The second-to-last row total shows how many drugs were detected via the ADR, and the last row total shows how many drugs were only detected via the ADR.

Multimedia Appendix 1 indicates that nervous system disorder and gastrointestinal disorder are the 2 ADRs that contributed the most number of repositioning drugs, while the ADR carpal tunnel syndrome suggested no repositioning drug. Among the repositioning drugs that have evidence support, 54% (15/28) were detected via only 1 ADR and 46% (13/28) were detected via more than 1 ADR. Moreover, within the 13 drugs, 6 of them were detected via the ADR combination: nervous system disorder + gastrointestinal disorder, which suggests that nervous system disorder and gastrointestinal disorder might be a potential ADR combination for identifying repositioning drugs for PD. In addition, to investigate whether there is relationship between the number of ADRs and the number of PubMed sources, we did correlation analysis. The calculated correlation coefficient (Pearson *R*) approximately equals 0, indicating little correlation between those 2 factors.

Analysis of the Repositioning Drugs Based on Literature

Table 2 summarizes the drugs analyzed in this section and their original labeled indications.

Table 1. Repositioning drugs and literature evidence.

Drug	Positive evidence	Negative evidence	No evidence
amphotericin B	✓	✓	
bupropion	✓		
carbamazepine	✓		
citalopram	✓	✓	
clomipramine	✓		
diltiazem		✓	
donepezil	✓		
fluoxetine	✓		
fluvoxamine	✓		
gabapentin	✓	✓	
glatiramer acetate	✓		
lamotrigine	✓	✓	
levetiracetam	✓		
methylphenidate	✓		
mirtazapine	✓		
modafinil	✓		
nefazodone	✓		
oxcarbazepine	✓		
paroxetine	✓		
phenytoin	✓		
rivastigmine	✓		
salbutamol	✓		
sertraline	✓		
thalidomide	✓		
topiramate	✓		
tramadol		✓	
valproic acid	✓		
ziprasidone	✓	✓	
alprazolam			✓
budesonide			✓
clarithromycin			✓
granisetron			✓
lansoprazole			✓
levonorgestrel			✓
mefloquine			✓
mycophenolic acid			✓
omeprazole			✓
tacrolimus			✓
tiagabine			✓
tobramycin			✓
topotecan			✓
valganciclovir			✓

Drug	Positive evidence	Negative evidence	No evidence
venlafaxine			✓
vigabatrin			✓

Table 2. Repositioning drugs and their labeled indications.

Labeled indication	Drug name
Immunomodulatory	Thalidomide
Multiple sclerosis	Glatiramer acetate
Asthma	Salbutamol
Depression	Bupropion, citalopram, clomipramine, paroxetine, sertraline, fluoxetine, fluvoxamine, mirtazapine, nefazodone
Alzheimer disease	Donepezil, rivastigmine
Epilepsy	Carbamazepine, gabapentin, lamotrigine, levetiracetam, topiramate, valproic acid, oxcarbazepine, phenytoin
Fatigue	Methylphenidate, modafinil

Thalidomide

Thalidomide is an immunomodulatory drug used mainly for certain cancers. The studies using mice showed that thalidomide improved the neurotoxicity induced by 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine as seen by a significant increase of dopamine, suggesting that thalidomide can improve the functional damage of the nigrostriatal cell substratum by the production of dopamine [20]. The neuroprotective effect makes thalidomide a potential adjuvant medication for PD. In addition, lenalidomide, a thalidomide derivative, was found to reduce motor behavioral deficits and improve dopaminergic fiber loss in the striatum by reducing microgliosis in both the striatum and hippocampus. This finding supports the potential of lenalidomide to address maladaptive neuroinflammation in PD [21].

Glatiramer Acetate

Conventional treatments for PD mainly address the dopamine deficiency, but some postmortem studies have showed that brain-derived neurotrophic factor (BDNF) deficiency may also play a role in PD pathogenesis, and this is supported by the finding that BDNF therapy is effective in animal models of PD [22]. Glatiramer acetate is an immunotherapy drug approved for treating multiple sclerosis. As glatiramer acetate can enhance central BDNF activity and augment neurogenesis, it may be useful to address the BDNF deficiency in PD; by exerting an anti-inflammatory effect, it can address the inflammatory process in the brain associated with PD [23]. According to this evidence, glatiramer acetate can be a potential medication for PD.

Salbutamol

Salbutamol is mainly used for asthma, bronchitis, emphysema, or other bronchospasms in lung diseases. There are preliminary studies showing that PD patients who were given salbutamol as adjunctive therapy improved in response to levodopa, as salbutamol enhances transport of levodopa across the blood-brain barrier [24]. An open-label pilot study also suggested that muscle mass and therapeutic response to levodopa in PD patients with fluctuation were improved by salbutamol [25]. To confirm its effects on PD, further studies such as double-blind and placebo-controlled studies are needed.

Antidepressants

Beside the motor dysfunctions in PD, the occurrence of nonmotor symptoms, such as psychological disorders, is very high. Specifically, depression occurs in 20% to 50% of PD patients, associated with increasing disability. Some antidepressants have been used in PD, but few studies have been conducted to support their efficacy and investigate adverse effects for PD patients. Therefore, an adequate therapeutic answer for treating depression together with PD is needed. Among the identified repositioning drugs, several are antidepressants.

Bupropion

Dopamine agonists have been considered as antidepressants in some studies, but they may cause side effects such as confusion, somnolence, and dizziness, so the role of the dopamine agonist in depressive PD still needs to be explored [26]. Bupropion, as a dopaminergic and noradrenergic antidepressant, could be a possible treatment for depressive symptoms associated with PD without serotonin-related side effects. Moreover, depression was improved in 5 clinical reports when using bupropion on PD patients while treatments with several other antidepressants were unsuccessful [27]. For instance, panic disorder was improved markedly on a 57-year-old female PD patient after introducing bupropion [28].

Citalopram

Citalopram is an antidepressant belonging to a group of drugs called selective serotonin reuptake inhibitors. As recent research indicates that selective noradrenergic (atomoxetine) and serotonergic (citalopram) reuptake inhibitors show potential to improve response inhibition on some PD patients, a double-blind experiment was conducted to investigate the behavioral efficacy of citalopram and atomoxetine [29]. Results supported the hypothesis that they are effective in inhibitory control. However, a retrospective survey showed that citalopram may trigger acute dystonia and exacerbate the abnormal movements of PD [30].

Therefore, the efficacy or adverse effect of citalopram for PD needs to be explored further.

Clomipramine

Clomipramine is a tricyclic antidepressant used to treat obsessive-compulsive disorder by increasing the activities of certain chemicals in brain. It was reported that delusions and hallucination conditions were improved by using clomipramine on a PD patient with depression [31].

Paroxetine

Through a 12-week placebo-controlled clinical experiment, paroxetine was found to improve the affective symptoms, somatic symptoms, and cognitive symptoms in depressed PD patients [32].

We did a literature analysis on the other 5 antidepressants (sertraline, fluoxetine, fluvoxamine, mirtazapine, and nefazodone) as well and found that those drugs are effective in dealing with the depression of PD patients. As the studies mostly discuss the effectiveness of the antidepressants on depression of PD patients rather than on PD itself only, we did not present all of the details here.

Medications for Alzheimer Disease

Donepezil

Donepezil, a cholinesterase inhibitor that works by increasing the amount of acetylcholine in the brain to reduce dementia symptoms, is indicated for Alzheimer disease. An exploratory study of 9 patients was conducted to study the effects of donepezil in PD patients with dementia, especially the effects of dose escalation [33], and the results showed that certain dose escalations of donepezil are useful for patients in the long term. In addition, the results of a randomized double-blind study showed that the combined use of donepezil and Di-Huang-Yi-Zhi is very effective in treating PD dementia, with a possible underlying reason that problems with the cholinergic system were ameliorated by such a combination therapy [34].

Rivastigmine

Rivastigmine is a drug used to treat Alzheimer disease and proven to be effective for mild cognitive impairment in PD. A double-blind placebo-controlled clinical study showed that a global rating of cognition, health status, and anxiety severity as well as the cognitive abilities according to a performance-based measure were improved by using rivastigmine [35].

Antiepilepsy Medications

Gabapentin

Gabapentin is an antiepileptic drug that affects the chemicals and nerves that cause seizures and pain to treat epilepsy and some types of nerve pain. Common medications for treating visual hallucination and the decrease of dopamine agonists often cause an exacerbation of motor symptoms, while gabapentin can directly affect the glutamic acid neuron system and the γ-amino butyric acid neuron system as an antiepilepsy drug. In a case report, both visual hallucination and pain of a PD patient were alleviated by using gabapentin, without any adverse effects [36]. However, in several cases, dyskinesia and bilateral ballism were induced when using gabapentin in PD diseases. Further studies are needed to prove the efficacy of this drug.

Lamotrigine

A group of researchers did both in vivo and in vitro experiments to examine whether safinamide (monoamine oxidase B and sodium channel blocker) has effects on microglial activation and the degeneration of dopaminergic neurons, which are closely associated with PD [37]. Results showed that safinamide has positive effects on microglial activation and protects dopaminergic neurons from degeneration in the 6-hydroxydopamine model of PD. In the experiments, rasagiline, a monoamine oxidase B inhibitor, and lamotrigine, a sodium channel–blocking drug, both exerted the protection of dopaminergic neurons, suggesting that safinamide can function in either or both mechanisms. In addition, the usefulness of lamotrigine in PD has been proven in several other studies, including a double-blind study on 20 patients [38] and an in vivo experiment on mice [39].

Levetiracetam

Levetiracetam was proved to be useful for the levodopa-induced dyskinesias in PD in a double-blind, placebo-controlled crossover trial [40].

Topiramate

Topiramate is mainly used to treat seizures in certain patients and sometimes used to treat migraine headaches. As it reduced levodopa-induced dyskinesia in animal models without affecting PD symptoms, a randomized, double-blind, placebo-controlled trial was then conducted on 13 patients. However, the results showed that topiramate worsened dyskinesia in PD patients [41].

Valproic Acid

Valproic acid is used for treating epilepsy by affecting chemicals that cause seizures [42]. An in vitro model was applied to investigate the treatment of valproic acid on synaptic loss and the underlying molecular mechanism. The experiment showed that the synaptic damage induced by amyloid-β, which is associated with Alzheimer disease, and by another neurodegenerative-associated protein, α-synuclein, which is associated with PD, could be reduced and protected by valproic acid [43], as valproic acid can inhibit the aberrant activation of amyloid-β–dependent cytoplasmic phospholipase A2. Therefore, valproic acid may be a potential therapy for Alzheimer disease and PD [44].

The other 2 antiepilepsy drugs that were found to be useful for PD are oxcarbazepine and phenytoin. Studies were conducted to evaluate the effectiveness of using antiepileptic drugs (eg, oxcarbazepine) together with some 3-hydroxy-3-methylglutaryl–coenzyme A reductase inhibitors for prevention of PD and other neurological diseases, and the experiments showed that the dopaminergic effect of oxcarbazepine is useful in the treatment of PD [45]. Another study, however, suggests that oxcarbazepine should be used with care because of its possible psychiatric side effects [46]. Epilepsy and PD often co-occur in the elderly, and phenytoin is used to control certain types of epilepsy, so many PD patients take phenytoin. Studies have been conducted to observe PD patients taking phenytoin, and a relationship is suspected between the function of phenytoin and the pathogenesis of PD [47].

Drugs Originally Used for Fatigue

Methylphenidate

Methylphenidate is a central nervous system stimulant used for hyperactivity disorder, found to be useful to address the maladaptive behavioral or cognitive aspects of fatigue in Parkinson patients [48].

Modafinil

Modafinil is a wakefulness-promoting agent used for sleep disorder and an effective treatment for fatigue in PD [49].

Other Drugs Known to Cause Adverse Effects in Parkinson Patients

Amphotericin B

Amphotericin B is used for treating progressive and potentially life-threatening fungal infections, but it was found to have interactions with PD in some clinical reports. There are several reports of neurotoxicity being caused by the increasing use of intrathecal amphotericin B—for instance, transient signs of parkinsonism occurred in 1 patient receiving the drug for cryptococcal meningitis [50] and in 3 children receiving it for pulmonary aspergillosis or sinus aspergillosis [51]. Although the indications of amphotericin B do not involve the nervous system, it was thought to have a direct toxic effect on the nervous tissue.

Diltiazem

Diltiazem is a calcium channel blocker that can relax the muscles of heart and blood vessels and is mainly used for hypertension and angina. However, it was found to induce Parkinson syndrome in some patients [52-53].

Tramadol

Tramadol is an analgesic widely prescribed because of its low abuse potential, but researchers found serotonin syndrome, a life-threatening adverse reaction, in a PD patient who used tramadol and ziprasidone [54].

Ziprasidone

Ziprasidone belongs to the class of atypical antipsychotics and is used for central or mood disorders. A clinical trial that conducted parallel comparison experiments on 14 patients demonstrated that ziprasidone is effective in ameliorating psychotic symptoms in PD patients [55]. However, serotonin syndrome occurred in a PD patient using ziprasidone for bipolar disorder [54]. Therefore, further studies are needed to evaluate the efficacy of ziprasidone for PD.

Discussion

Principal Findings

In drug repositioning research, results are only suggestions or predictions rather than confirmed truth until clinical trials are conducted. As a result, the evaluation of method performance becomes a difficult task, because there is no gold standard. In previous studies, there have been two common ways of evaluation: computational assessment, based on the co-occurrence of drug-disease terms in biomedical literature, and experimental assessment, based on in silico or in vitro experiments. Here we have adopted the computational assessment but beyond counting the co-occurrence frequency, we conducted a literature review to investigate whether the scientific studies in the published articles supported the repositioning of the drug to PD. In the future, experimental assessment could be exploited to reinforce the evaluation.

Among the discovered repositioning drugs, some have the potential to treat PD via their neuroprotection or neurotoxicity prevention function, while several others show the effects in clinical practice without the underlying mechanism being known; therefore, exhaustive analysis is needed of the drugs. In addition, we identified several psychological drugs that are proven to be effective without inducing adverse effects when used to treat PD. As psychological symptoms occur in a high ratio of PD patients, these findings are quite useful. Based on the results, the repositioning drugs that have positive functions on the neural system are the most promising findings and are the primary candidates for further experimental assessment. Meanwhile, the repositioning findings that are commonly used for other complicating diseases on PD patients should be tested more carefully for efficacy and risk evaluation.

In this work, we relied on the rational association between disease and ADR to find potential repositioning drugs and achieved an appreciable result. To investigate whether there is any relationship between the number of intermediary ADRs and the number of supporting literature sources, we did a correlation analysis, but there is little correlation between the 2 factors according to the calculated Pearson *R* coefficient (≈0). A possible reason could be the size of our dataset (eg, ADR and drug). In a future study, we can expand the dataset size to include more ADRs and involve more biomedical characteristics, such as chemical or biological characteristics or a combination of the two, hoping for better performance.

Conclusions

In this study, we proposed a computational drug repositioning method based on a heterogeneous network and used the rational association between disease and ADR to reveal repositioning drugs. Experiment results demonstrated that our approach is able to identify FDA-approved novel drugs for PD, and many of the predictions are supported by existing studies according to the literature-based investigation. Although further studies are needed to confirm the pharmaceutical effects of these repositioning drugs for PD, this study suggests the possibility and effectiveness of applying systematic drug repositioning methods to drug development.

Acknowledgments

This work was supported in part by the National Science Foundation under grants NSF-1741306, NSF-1650531, and NSF-1443019.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Repositioning drugs detected via each adverse drug reaction.

[PDF File (Adobe PDF File), 46KB - jmir_v20i10e271_app1.pdf]

References

1. Warner T, Schapira A. Genetic and environmental factors in the cause of Parkinson's disease. Ann Neurol 2003;53 Suppl 3:S16-S23. [doi: 10.1002/ana.10487] [Medline: 12666095]
2. Dudley JT, Deshpande T, Butte AJ. Exploiting drug-disease relationships for computational drug repositioning. Brief Bioinform 2011 Jul;12(4):303-311 [FREE Full text] [doi: 10.1093/bib/bbr013] [Medline: 21690101]
3. Zhang P, Wang F, Hu J. Towards drug repositioning: a unified computational framework for integrating multiple aspects of drug similarity and disease similarity. AMIA Annu Symp Proc 2014;2014:1258-1267 [FREE Full text] [Medline: 25954437]
4. Zhang P, Wang F, Hu J. Exploring the relationship between drug side-effects and therapeutic indications. 2013 Presented at: AMIA Annual Symposium Proceedings. American Medical Informatics Association; 2013; Washington p. 1568.
5. Yang L, Agarwal P. Systematic drug repositioning based on clinical side-effects. PLoS One 2011;6(12):e28025 [FREE Full text] [doi: 10.1371/journal.pone.0028025] [Medline: 22205936]
6. Yang H, Yang CC. Discovering drug-drug interactions and associated adverse drug reactions with triad prediction in heterogeneous healthcare networks. 2016 Presented at: Healthcare Informatics (ICHI), 2016 IEEE International Conference; 2016; Chicago p. 244-254.
7. Yang CC, Yang H. Exploiting social media with tensor decomposition for pharmacovigilance. 2015 Presented at: Data Mining Workshop (ICDMW), 2015 IEEE International Conference; 2015 Nov 14; Atlantic City p. 188-195.
8. Yang C, Yang H, Jiang L. Postmarketing drug safety surveillance using publicly available health-consumer-contributed content in social media. ACM Transact Manag Inf Syst 2014;5(1):2. [doi: 10.1145/2576233]
9. Yang H, Yang CC. Using health-consumer-contributed data to detect adverse drug reactions by association mining with temporal analysis. ACM Transactions on Intelligent Systems and Technology 2015;6(4):55. [doi: 10.1145/2700482]
10. Ng C, Hauptman R, Zhang Y, Bourne PE, Xie L. Anti-infectious drug repurposing using an integrated chemical genomics and structural systems biology approach. Pac Symp Biocomput 2014:136-147 [FREE Full text] [Medline: 24297541]
11. Wang W, Yang S, Zhang X, Li J. Drug repositioning by integrating target information through a heterogeneous network model. Bioinformatics 2014 Oct 15;30(20):2923-2930 [FREE Full text] [doi: 10.1093/bioinformatics/btu403] [Medline: 24974205]
12. Wu C, Gudivada RC, Aronow BJ, Jegga AG. Computational drug repositioning through heterogeneous network clustering. BMC Syst Biol 2013;7 Suppl 5:S6 [FREE Full text] [doi: 10.1186/1752-0509-7-S5-S6] [Medline: 24564976]
13. Zeng QT, Tse T. Exploring and developing consumer health vocabularies. J Am Med Inform Assoc 2006;13(1):24-29 [FREE Full text] [doi: 10.1197/jamia.M1761] [Medline: 16221948]
14. Jiang L, Yang C. Expanding consumer health vocabularies by learning consumer health expressions from online health social media. 2015 Presented at: International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction; 2015; Washington p. 314-320.
15. William M, Dagli CK. Social network analysis with content and graphs. Lincoln Lab J 2013:61-81.
16. Sun Y, Han J. Mining heterogeneous information networks: principles and methodologies. Syn Lect Data Min Knowl Disc 2012;3(2):1-159. [doi: 10.2200/S00433ED1V01Y201207DMK005]
17. Kuhn M, Letunic I, Jensen LJ, Bork P. The SIDER database of drugs and side effects. Nucleic Acids Res 2015 Oct 19;44(D1):D1075-D1079 [FREE Full text] [doi: 10.1093/nar/gkv1075] [Medline: 26481350]
18. Whirl-Carrillo M, McDonagh E, Hebert J, Gong L, Sangkuhl K, Thorn C, et al. Pharmacogenomics knowledge for personalized medicine. Clin Pharmacol Ther 2012 Oct;92(4):414-417 [FREE Full text] [doi: 10.1038/clpt.2012.96] [Medline: 22992668]
19. MedHelp. 2016. URL: http://www.medhelp.org/ [accessed 2016-10-10] [WebCite Cache ID 71kEYg9wd]
20. Palencia G, Garcia E, Osorio-Rico L, Trejo-Solís C, Escamilla-Ramírez A, Sotelo J. Neuroprotective effect of thalidomide on MPTP-induced toxicity. Neurotoxicology 2015 Mar;47:82-87. [doi: 10.1016/j.neuro.2015.02.004] [Medline: 25724264]
21. Valera E, Mante M, Anderson S, Rockenstein E, Masliah E. Lenalidomide reduces microglial activation and behavioral deficits in a transgenic model of Parkinson's disease. J Neuroinflamm 2015 May 14;12:1 [FREE Full text] [doi: 10.1186/s12974-015-0320-x] [Medline: 25966683]

22. Tsai S. Glatiramer acetate could be a potential therapeutic agent for Parkinson's disease through its neuroprotective and anti-inflammatory effects. Med Hypotheses 2007;69(6):1219-1221. [doi: 10.1016/j.mehy.2007.04.014] [Medline: 17548170]
23. Lalive P, Neuhaus O, Benkhoucha M, Burger D, Hohlfeld R, Zamvil S, et al. Glatiramer acetate in the treatment of multiple sclerosis: emerging concepts regarding its mechanism of action. CNS Drugs 2011 May;25(5):401-414 [FREE Full text] [doi: 10.2165/11588120-000000000-00000] [Medline: 21476611]
24. Uc EY. Salbutamol potentiates effect of levodopa in Parkinson's disease. Inpharma 2003:27.
25. Uc EY, Lambert CP, Harik SI, Rodnitzky RL, Evans WJ. Albuterol improves response to levodopa and increases skeletal muscle mass in patients with fluctuating Parkinson disease. Clin Neuropharmacol 2003;26(4):207-212. [Medline: 12897642]
26. Raskin S, Durst R. Bupropion as the treatment of choice in depression associated with Parkinson's disease and its various treatments. Med Hypotheses 2010 Dec;75(6):544-546. [doi: 10.1016/j.mehy.2010.07.024] [Medline: 20708340]
27. Załuska M, Dyduch A. Bupropion in the treatment of depression in Parkinson's disease. Int Psychogeriatr 2011 Mar;23(2):325-327. [doi: 10.1017/S1041610210001687] [Medline: 20809995]
28. Gebhardt S, Röttgers H, Bäcker A. Treatment of panic disorder with bupropion in a patient with Parkinson's disease. J Clin Pharm Ther 2008;33(5):575-577. [Medline: 18834374]
29. Ye Z, Rae C, Nombela C. Predicting beneficial effects of atomoxetine and citalopram on response inhibition in Parkinson's disease with clinical and neuroimaging measures. Human Brain Mapping 2016;37(3):1026-1037. [Medline: 26757216]
30. Moosavi S, Ahmadi M, Monajemi M. Acute dystonia due to citalopram treatment: a case series. Glob J Health Sci 2014 Oct 28;6(6):295-299 [FREE Full text] [doi: 10.5539/gjhs.v6n6p295] [Medline: 25363102]
31. Meco G, Bernardi S. Antidepressant use in treatment of psychosis with comorbid depression in Parkinson's disease. Prog Neuropsychopharmacol Biol Psychiatry 2007 Jan 30;31(1):311-313. [doi: 10.1016/j.pnpbp.2006.06.016] [Medline: 16919377]
32. Broen M, Leentjens A, Köhler S, Kuijf M, McDonald W, Richard I. Trajectories of recovery in depressed Parkinson's disease patients treated with paroxetine or venlafaxine. Parkinsonism Relat Disord 2016 Feb;23:80-85. [doi: 10.1016/j.parkreldis.2015.12.013] [Medline: 26739248]
33. Ishikawa K, Motoi Y, Mizuno Y, Kubo S, Hattori N. Effects of donepezil dose escalation in Parkinson's patients with dementia receiving long-term donepezil treatment: an exploratory study. Psychogeriatrics 2014 Jun;14(2):93-100 [FREE Full text] [doi: 10.1111/psyg.12045] [Medline: 24661498]
34. Gu C, Shen T, An H, Yuan C, Zhou J, Ye Q, et al. Combined therapy of Di-Huang-Yi-Zhi with Donepezil in patients with Parkinson's disease dementia. Neurosci Lett 2015 Oct 08;606:13-17. [doi: 10.1016/j.neulet.2015.08.019] [Medline: 26297124]
35. Onor M, Trevisiol M, Aguglia E. Rivastigmine in the treatment of Alzheimer's disease: an update. Clin Interv Aging 2007;2(1):17-32 [FREE Full text] [Medline: 18044073]
36. Abe K, Chiba Y, Katsuse O, Hirayasu Y. A case of Parkinson disease with both visual hallucination and pain improved by gabapentin. Clin Neuropharmacol 2016;39(1):55-56. [doi: 10.1097/WNF.0000000000000122] [Medline: 26757314]
37. Sadeghian M, Mullali G, Pocock J, Piers T, Roach A, Smith K. Neuroprotection by safinamide in the 6-hydroxydopamine model of Parkinson's disease. Neuropathol Appl Neurobiol 2016 Dec;42(5):423-435. [doi: 10.1111/nan.12263] [Medline: 26300398]
38. Zipp F, Bürklin F, Stecker K, Baas H, Fischer P. Lamotrigine in Parkinson's disease—a double blind study. J Neural Transm Park Dis Dement Sect 1995;10(2-3):199-206. [Medline: 9620063]
39. Jones-Humble S, Morgan P, Cooper B. The novel anticonvulsant lamotrigine prevents dopamine depletion in C57 black mice in the MPTP animal model of Parkinson's disease. Life Sci 1994;54(4):245-252. [Medline: 8289584]
40. Stathis P, Konitsiotis S, Tagaris G, Peterson D, VALID-PD Study Group. Levetiracetam for the management of levodopa-induced dyskinesias in Parkinson's disease. Mov Disord 2011 Feb 01;26(2):264-270. [doi: 10.1002/mds.23355] [Medline: 21412833]
41. Kobylecki C, Burn D, Kass-Iliyya L, Kellett M, Crossman A, Silverdale M. Randomized clinical trial of topiramate for levodopa-induced dyskinesia in Parkinson's disease. Parkinsonism Relat Disord 2014 Apr;20(4):452-455. [doi: 10.1016/j.parkreldis.2014.01.016] [Medline: 24521874]
42. Sugai F, Yamamoto Y, Miyaguchi K, Zhou Z, Sumi H, Hamasaki T, et al. Benefit of valproic acid in suppressing disease progression of ALS model mice. Eur J Neurosci 2004 Dec;20(11):3179-3183. [doi: 10.1111/j.1460-9568.2004.03765.x] [Medline: 15579172]
43. Choonara Y, Pillay V, du T, Modi G, Naidoo D, Ndesendo V, et al. Trends in the molecular pathogenesis and clinical therapeutics of common neurodegenerative disorders. Int J Mol Sci 2009 Jun 03;10(6):2510-2557 [FREE Full text] [doi: 10.3390/ijms10062510] [Medline: 19582217]
44. Williams R, Bate C. An in vitro model for synaptic loss in neurodegenerative diseases suggests a neuroprotective role for valproic acid via inhibition of cPLA2 dependent signalling. Neuropharmacology 2016 Feb;101:566-575. [doi: 10.1016/j.neuropharm.2015.06.013] [Medline: 26116815]
45. Okuyucu E, Duman T, Akcin E. Reversible parkinsonism with oxcarbazepine use. Parkinsonism Relat Disord 2009 Dec;15(10):787-788. [doi: 10.1016/j.parkreldis.2009.03.007] [Medline: 19401271]
46. Kovacs N, Nagy F, Balas I, Komoly S, Janszky J. Oxcarbazepine may induce psychotic symptoms in Parkinson's disease. Epilepsy Behav 2008 Apr;12(3):492-493. [doi: 10.1016/j.yebeh.2007.12.010] [Medline: 18222110]

47. Ferrari M, Peeters E, Haan J, Roos R, Vermey P, de Wolff F, et al. Cytochrome P450 and Parkinson's disease. Poor parahydroxylation of phenytoin. J Neurol Sci 1990 May;96(2-3):153-157. [Medline: 2376747]
48. Moreau C, Delval A, Defebvre L, Dujardin K, Duhamel A, Petyt G, Parkgait-II study group. Methylphenidate for gait hypokinesia and freezing in patients with Parkinson's disease undergoing subthalamic stimulation: a multicentre, parallel, randomised, placebo-controlled trial. Lancet Neurol 2012 Jul;11(7):589-596. [doi: 10.1016/S1474-4422(12)70106-0] [Medline: 22658702]
49. Sheng P, Hou L, Wang X, Wang X, Huang C, Yu M, et al. Efficacy of modafinil on fatigue and excessive daytime sleepiness associated with neurological disorders: a systematic review and meta-analysis. PLoS One 2013;8(12):e81802 [FREE Full text] [doi: 10.1371/journal.pone.0081802] [Medline: 24312590]
50. Fisher JF, Dewald J. Parkinsonism associated with intraventricular amphotericin B. J Antimicrob Chemother 1983 Jul;12(1):97-99. [Medline: 6619049]
51. Mott SH, Packer RJ, Vezina LG, Kapur S, Dinndorf PA, Conry JA, et al. Encephalopathy with parkinsonian features in children following bone marrow transplantations and high-dose amphotericin B. Ann Neurol 1995 Jun;37(6):810-814. [doi: 10.1002/ana.410370616] [Medline: 7778856]
52. Remblier C, Kassir A, Richard D, Pérault MC, Guibert S. [Parkinson syndrome from diltiazem]. Therapie 2001;56(1):57-59. [Medline: 11322019]
53. Graham DF, Stewart-Wynne EG. Diltiazem-induced acute parkinsonism. Aust N Z J Med 1994 Feb;24(1):70. [Medline: 8002867]
54. El-Okdi N, Lumbrezer D, Karanovic D, Ghose A, Assaly R. Serotonin syndrome after the use of tramadol and ziprasidone in a patient with a deep brain stimulator for Parkinson disease. Am J Ther 2014;21(4):e97-e99. [doi: 10.1097/MJT.0b013e3182456d88] [Medline: 24158007]
55. Pintor L, Valldeoriola F, Baillés E, Martí MJ, Muñiz A, Tolosa E. Ziprasidone versus clozapine in the treatment of psychotic symptoms in Parkinson disease: a randomized open clinical trial. Clin Neuropharmacol 2012;35(2):61-66. [doi: 10.1097/WNF.0b013e31824d5115] [Medline: 22388466]

Abbreviations

ADR: adverse drug reaction
BDNF: brain-derived neurotrophic factor
CHV: Consumer Health Vocabulary
PD: Parkinson disease
PharmGKB: Pharmacogenomics Knowledgebase
SIDER: Side Effect Resource database

Edited by G Eysenbach; submitted 13.12.17; peer-reviewed by J Zhou, R Haase; comments to author 22.03.18; revised version received 17.05.18; accepted 18.06.18; published 11.10.18.

Please cite as:
Zhao M, Yang CC
Drug Repositioning to Accelerate Drug Development Using Social Media Data: Computational Study on Parkinson Disease
J Med Internet Res 2018;20(10):e271
URL: http://www.jmir.org/2018/10/e271/
doi:10.2196/jmir.9646
PMID:30309833

Original Paper

Nature and Diffusion of Gynecologic Cancer–Related Misinformation on Social Media: Analysis of Tweets

Liang Chen[1*], PhD; Xiaohui Wang[2*], PhD; Tai-Quan Peng[3], PhD

[1]Lab for Big Data and Public Communication, School of Communication and Design, Sun Yat-sen University, Guangzhou, China
[2]Department of Journalism, School of Communication, Hong Kong Baptist University, Hong Kong, China (Hong Kong)
[3]Department of Communication, Michigan State University, East Lansing, MI, United States
[*]these authors contributed equally

Corresponding Author:
Tai-Quan Peng, PhD
Department of Communication
Michigan State University
404 Wilson Road, Room 473
East Lansing, MI, 48824
United States
Phone: 1 5173550221
Email: pengtaiq@msu.edu

Abstract

Background: Over the last two decades, the incidence and mortality rates of gynecologic cancers have increased at a constant rate in China. Gynecologic cancers have become one of the most serious threats to women's health in China. With the widespread use of social media, an increasing number of individuals have employed social media to produce, seek, and share cancer-related information. However, health information on social media is not always accurate. Health, and especially cancer-related, misinformation has been widely spread on social media, which can affect individuals' attitudinal and behavioral responses to cancer.

Objective: The aim of this study was to examine the nature and diffusion of gynecologic cancer–related misinformation on Weibo, the Chinese equivalent of Twitter.

Methods: A total of 2691 tweets related to 2 gynecologic cancers—breast cancer and cervical cancer—posted on Weibo from June 2015 to June 2016 were extracted using the Python Web Crawler. Two medical school graduate students with expertise in gynecologic diseases were recruited to code the tweets to differentiate between true information and misinformation as well as to identify the types of falsehoods. The diffusion characteristics of gynecologic cancer–related misinformation were compared with those of the true information.

Results: While most of the gynecologic cancer–related tweets provided medically accurate information, approximately 30% of them were found to contain misinformation. Furthermore, it was found that tweets about cancer treatment contained a higher percentage of misinformation than prevention-related tweets. Nevertheless, the prevention-related misinformation diffused significantly more broadly and deeply than true information on social media.

Conclusions: The findings of this study suggest the need for controlling and reducing the cancer-related misinformation on social media with the efforts from both service providers and medical professionals. More specifically, it is important to correct falsehoods related to the prevention of gynecologic cancers on social media and increase individuals' capacity to assess the veracity of Web-based information to curb the spread and thus minimize the consequences of cancer-related misinformation.

(J Med Internet Res 2018;20(10):e11515) doi:10.2196/11515

KEYWORDS

social media; breast cancer; cervical cancer; misinformation; diffusion; China

Introduction

Background

In recent years, cancer has become a major public health issue in China. According to the statistics from the National Central Cancer Registry, there were approximately 3.80 million new cancer cases and 2.30 million cancer-caused deaths in China in 2014 [1]. For women, 2 out of the 10 most common cancers are gynecologic cancers, with breast cancer (268,600 new cases) and cervical cancer (98,900 new cases) being the most prevalent in 2015 [2]. The incidence and mortality rates of gynecologic cancers have increased substantially in China over the last 2 decades, and it has been 1 of the major health concerns for women in China [2].

With the rapid development of social media, social media has become a popular means by which individuals can access a staggering amount of health information [3,4]. A growing number of individuals, especially women, turn to social media to seek out and share a variety of cancer-related information, such as seeking information about cancer prevention and treatment as well as sharing the experience of having it, and to obtain social support to cope with the disease and manage emotions [5,6]. Medical professionals and traditional portals contribute to the health information available on social media, but a greater amount of information is generated and disseminated by ordinary users based on their first-hand cancer experiences [7]. Web-based health information has been found to be effective in raising individuals' awareness of diseases and fueling communication between lay persons and health care professionals [8]. Furthermore, Web-based health information could help individuals improve their abilities to prevent certain diseases and enable them to effectively manage chronic health conditions [9,10].

Nevertheless, individuals might take great risks in their utilization of Web-based resources, as health information on social media is not always accurate [11,12]. It has been shown that health-related misinformation in general, and cancer-related misinformation in particular, has been widely spread on social media, which affects individuals' responses to cancer prevention and treatment [13]. Moreover, because of overloaded information on social media, ordinary users may not have the resources, knowledge, and expertise to assess the veracity of Web-based cancer-related information and to identify informative and trustworthy information on social media [13].

Although scholarly attention has been drawn to misinformation on social media, little is known about the nature and diffusion of cancer-related misinformation there. In this study, we filled this gap by examining the nature and diffusion of misinformation about breast cancer and cervical cancer on social media. Specifically, we used a content analysis not only to differentiate between true information and misinformation regarding these 2 types of gynecologic cancers on social media but also to identify the types of falsehoods in such information. Furthermore, the diffusion characteristics of cancer-related misinformation were examined and compared with those of true information.

Misinformation on Social Media

Misinformation refers to false and inaccurate information that is spread intentionally or unintentionally [14]. With the increasing penetration of information and communication technologies, massive amounts of misinformation can be easily disseminated to a larger group of audience at very low costs. One study revealed that on average, individuals in the United States encountered 1 to 3 fake news stories online in the month before the 2016 US election [15]. About one-quarter of adults reported having shared fabricated political news online, sometimes by mistake and sometimes intentionally [16].

Many scholars have argued that social media is responsible for the high prevalence of Web-based misinformation [17]. Traditional media content is usually produced by professional journalists and editors who are information gatekeepers with adequate knowledge and resources to assess the veracity of information. However, ordinary users of social media are empowered to produce and share a wide variety of information irrespective of its veracity [18]. Thus, it is not surprising that there is a large amount of misinformation on social media.

The proliferation of Web-based misinformation has caused negative consequences for both individuals and the society as a whole. Specifically, misinformation, such as fake news, rumors, and inaccurate information, not only causes the spread of unnecessary fears and conspiracies but also distorts individuals' behavioral responses to certain issues, such as political elections, natural disasters, and diseases [16,19]. For example, misinformation about vaccinations makes many parents refuse immunizations for their children, which has led to a noticeable increase in vaccine-preventable diseases and has even caused deaths among children [20]. Furthermore, misinformation exerts negative impacts on our society that may trigger financial panic and even strain diplomatic relations [21]. The 2013 World Economic Forum listed misinformation as one of the main threats to human society [17].

To constrain the amount of misinformation spread on social media and to minimize the negative effects caused by the misinformation, many researchers have attempted to examine how misinformation spreads on social media and investigate the driving mechanisms that underlie the diffusion of Web-based misinformation in various domains, such as natural disasters, science, and politics. Specifically, Oh et al [22] analyzed the working dynamics of rumors related to the Haiti earthquake in 2010 based on data from Twitter and found that informational uncertainty and anxiety are key factors that determine the rapid spread of a rumor. Moreover, they indicated that reliable information with credible sources could reduce levels of anxiety on Twitter, which in turn limits the spread of rumors. Domenico et al [23] explored the spread of a scientific rumor about the Higgs boson and proposed a model for its spread. They found that individuals were more likely to spread the rumor if most of their friends tweet it repeatedly. More recently, Vosoughi et al [19] explored the diffusion structure of true and false news on Twitter and found that false news spread faster, deeper, and more broadly than the truth. Such differences in the diffusion of truth and falsehoods may be related to the fact that false news tends to include more emotion of fear, disgust, and surprise, which could cause the misinformation to go viral. Another explanation could be a novelty effect, that is, people are more willing to exchange novel information [19]. A study by Zhao et al [24] is one of the few studies that had investigated the misinformation on Chinese social media. They indicated that when large-scale social crises occur, a great number of rumors are posted and reposted quickly on social media. Moreover, they found that attitude and personal norms are the key factors, which would drive social media users in China to combat rumors.

Although much attention has been directed toward the spread of misinformation on social media, most studies have focused on political or scientific misinformation in general. Only a few studies have examined the nature and diffusion of health, especially cancer-related misinformation on social media. The uniqueness in the context of cancer-related information requires a specific investigation.

Cancer-Related Misinformation on Social Media

The rise of news media has created an atmosphere of hype and hysteria about cancer in which individuals have been exposed to conflicting information [25]. This leads to many misperceptions about cancer, including its causes, prevention, and treatment [25]. In the last decade, social media has exacerbated individuals' uncertainty about cancer. Unlike traditional media, most health-related content on social media is generated and shared by patients and caregivers based on their own personal experiences. The content may include many false elements that can distort individuals' attitudes and behaviors toward cancer prevention and treatment. Gage-Bouchard et al [26] conducted a content analysis to assess the veracity of information related to lymphoblastic leukemia on 19 public Facebook pages and found that at least one-third of the exchanged information was not medically or scientifically accurate.

With its extremely large population, China contributes significantly to the global burden of cancer [27]. Gynecologic cancers (eg, breast cancer, ovarian cancer, and cervical cancer) have become the most common cancers among Chinese women [2]. However, cancer is preventable if people are aware of its causes and science-based prevention strategies [28]. As more and more female users in China have used social media to exchange a variety of cancer-related information [6,26,29], this study focused on misinformation regarding breast cancer and cervical cancer on Chinese social media.

Research Questions

On the basis of the aforementioned literature, misinformation, especially health misinformation, has been prevalent on social media, which has drawn much attention from the governments, academia, and industry [13]. Recently, several studies have examined the veracity of cancer-related information on social media. However, most of these studies were exploratory in nature, only describing the prevalence of misinformation, and did not consider the types of misinformation and their diffusion characteristics [26]. To fill in the gaps in the literature, this study focuses on 2 gynecologic cancers, namely breast and cervical cancers, and proposes the following 3 research questions:

- Research question 1: What is the distribution of true information and misinformation regarding gynecologic cancers on Weibo, a Chinese version of Twitter?
- Research question 2: What kinds of gynecologic cancer–related misinformation exist on Weibo?
- Research question 3: How do the diffusion characteristics of true information and misinformation regarding gynecologic cancers differ?

To answer our research questions, a content analysis was first conducted to differentiate between true information and misinformation and to identify the types of falsehoods embedded in the misinformation. Next, the diffusion structure of each piece of cancer-related information on social media was constructed and analyzed with a network perspective to understand how misinformation was spread and received by social media users.

Methods

Data Collection

Two keywords "乳癌/乳腺癌" [breast cancer] and "子宫癌/宫颈癌" [cervical cancer] were employed to search tweets about breast cancer and cervical cancer on Weibo, one of the most popular social media platforms in China. We randomly selected 7 weeks out of 52 weeks from June 2015 to May 2016. Tweets posted in the 7 weeks were retrieved and included in the study. In total, 2691 tweets were extracted with the Python Web Crawler. The content, post time, and diffusion path of each tweet were retrieved.

In terms of ethical issues, Weibo is considered a public domain in which data are freely accessible to the public. To minimize the potential harm to Weibo users, all the data collected from Weibo were deindividualized to maintain the users' anonymity. Moreover, all of the tweets presented in this paper were paraphrased or written in aggregate to prevent identification of the users.

Coding Procedure

A total of 2 medical school graduate students with expertise in gynecologic diseases were recruited to complete the coding. Initially, the 2 coders were asked to pilot the project by coding 10.41% (280/2691) of the total tweets to develop and refine the coding schemes. Of the 2691 total tweets on Weibo, 1144 tweets (1144/2691, 42.51%) only expressed personal emotions and experiences that cannot be identified as truth or falsehood. These tweets were excluded from further analysis. The remaining 1547 (1547/2691, 57.49%) tweets contained medically oriented information for which the thematic category and information veracity were coded.

First, 4 *thematic categories* of all the 1547 tweets were coded: (1) background knowledge, which refers to basic information about breast cancer and cervical cancer, including the prevalence, causes, and symptoms of each cancer; (2) prevention, which refers to methods and actions that can lower the risk of getting the cancer under study, includes maintaining a healthy lifestyle, avoiding exposure to known cancer-causing substances, and taking medicines or vaccines; (3) diagnosis, which refers to the act of identifying a disease from its signs and symptoms; and (4) treatment, which refers to drugs or methods that can attack specific types of cancer cells to help the patient fight the disease. These 4 thematic categories are mutually exclusive, implying that each tweet will be assigned to 1 theme only. Krippendorff alpha [30] for this round coding was .95, which means that the intercoder reliability for thematic category is well accepted.

Second, the 2 coders coded the *information veracity* of the 1547 tweets. Specifically, the coders categorized each tweet as 1 (=true information) and 2 (=misinformation) for the tweets in the 4 thematic categories. Krippendorff alpha tests [30] revealed an acceptable level of intercoder reliability for all of the variables: .91 for background knowledge, .88 for prevention, .93 for diagnosis, and .89 for treatment.

Finally, when a tweet was categorized as *misinformation* in each thematic category, the 2 coders indicated the types of falsehoods using a conventional content analysis. Conventional content analysis is a qualitative approach widely used in health research [31]. First, all of the false tweets were read repeatedly to achieve immersion and obtain a sense of the entire situation [32]. Second, while reading the tweets, the 2 coders highlighted exact words from the text as codes or created new codes to capture key concepts [33]. Third, these codes were sorted into categories based on their relationships. Thereafter, the second and third procedures were repeated to keep the acceptability and reliability of the designated categories high. Finally, each category was defined. The validity of the coding was checked using a deviant case analysis.

Quantifying the Diffusion Characteristics of Gynecologic Cancer–Related Information

The diffusion characteristics of all the 1547 tweets were measured with following 5 characteristics: the scale of retweets, the range of retweets, the structural virality of retweets, the number of comments, and the number of likes. These 5 indices measure the diffusion breadth, diffusion depth, and the engagement of the information relevant to breast and cervical cancers [31].

The retweet network of all 1547 tweets was first constructed by tracking how each original tweet was retweeted. The scale, range, and structural virality of the retweet networks were estimated and assigned as diffusion indices for each tweet. The scale of retweets is the number of total retweets received by a tweet. The range of retweets refers to the depth of a retweet network as indicated by the number of hops in a diffusion chain [34]. The structural virality of retweets measures the divergent branches in the diffusion network [35], which is equal to the average distance between all pairs of nodes in a retweet network. Beyond the characteristics derived from the retweet networks, comments and likes received by a tweet can represent users' engagement in the process of information spreading [36]. Thus, the number of comments and the number of likes received by each tweet are included as the other 2 diffusion characteristics of the gynecologic cancer–related information in the study. The 5 characteristics capture the information diffusion on social media from a multidimensional perspective, which provides a more comprehensive understanding of the diffusion structures of Web-based information.

Results

Nature of Gynecologic Cancer–Related True Information and Misinformation

Among the 1547 medically oriented tweets, the most commonly exchanged type of cancer-related information was background knowledge (749/1547, 48.42%), followed by prevention (467/1547, 30.19%), treatment (189/1547, 12.21%), and diagnosis (142/1547, 9.18%). Moreover, 66.13% (1023/1547) of the tweets provided true information and 33.87% (524/1547) contained misinformation. A chi-square test indicated that true information was significantly more prevalent than misinformation, $\chi^2_{1(N=1547)}=160.9$, $P<.001$.

Information in 4 thematic categories was found to differ significantly on their information veracity, $\chi^2_{3(N=1547)}=322.5$, $P<.001$. Tweets on treatment contained a higher percentage of misinformation than true information; specifically, 156 (156/189, 82.5%) tweets related to cancer treatment included misinformation. These were followed by tweets about background knowledge; of these, 287 (287/749, 38.3%) tweets contained misinformation. Only 14.4% (67/467) of prevention-related tweets and 14 (14/142, 9.9%) diagnosis-related tweets were not medically accurate (see Table 1).

The types of falsehoods were identified for information in each thematic category as summarized in Table 2. Specifically, the falsehoods in the category of background knowledge mainly included epidemiology, risk factors, prognosis, and pathology. Prevention-related tweets had a relatively small amount of misinformation that involved 2 types of falsehoods: lifestyle and vaccinations. Diagnosis-related misinformation was divided into 2 types: clinical manifestations and diagnostic techniques. Cancer treatment–related misinformation mainly included surgery, radiation therapy, drug therapy, and other therapies.

Table 1. Distribution of gynecologic cancer–related information by thematic category and information veracity.

Information veracity	Thematic category			
	Background knowledge, n (%)	Prevention, n (%)	Diagnosis, n (%)	Treatment, n (%)
True information	462 (61.7)	400 (85.7)	128 (90.1)	33 (17.5)
Misinformation	287 (38.3)	67 (14.4)	14 (9.9)	156 (82.5)
Total	749 (100.0)	467 (100.0)	142 (100.0)	189 (100.0)

Table 2. Types of falsehoods in different thematic categories of gynecologic cancer–related information on Weibo.

Thematic category and types of falsehoods	Definition	Example
Background knowledge		
Epidemiology	The distribution and determinants of health and disease conditions in specified populations	The cancer prevalence rate is 10% higher in China than the world average
Risk factors	An aspect of personal behavior or lifestyle, environmental exposure, inborn or inherited characteristic, which on the basis of epidemiological evidence, is known to be associated with a health-related condition	Using preservative-containing cosmetics is one of the main causes of breast cancer
Pathology	A specialty concerned with the nature and cause of disease as expressed by changes in cellular or tissue structure and function caused by the disease process	Breast hyperplasia is the beginning of breast cancer
Prognosis	A prediction of the probable outcome of a disease based on an individual's condition and the usual course of the disease as seen in similar situations	Triple-negative breast cancer has a better prognosis than the normal type, and the 5-year survival rate is high
Prevention		
Lifestyle	Typical way of life or manner of living characteristic of an individual or group	Drinking 5 cups of coffee a day or regular exercise, such as cycling, could reduce the risk of developing breast cancer by at least 20%
Vaccinations	Administration of vaccines to stimulate individuals' immune responses	The HPV[a] vaccine can reduce the risk of cervical cancer by 100%
Diagnosis		
Clinical manifestations	A symptom is observed by the patient subjectively but cannot be measured directly, whereas a sign is objectively observable by others	Any abnormality of the breast is an early symptom of breast cancer
Diagnostic techniques and procedures	Methods, procedures, and tests performed to diagnose a disease, disordered function, or disability	A compound derived from urinary thiol is the only reagent that can detect early cervical cancer
Treatment		
Surgery	Operations conducted for the correction of deformities and defects, repair of injuries, and diagnosis and cure of certain diseases	Precancerous lesions in the endometrium indicate the need for surgery to remove the uterus
Radiotherapy	Ionizing radiation conducted to treat malignant neoplasms and some benign conditions	Up to 60% of cancer patients need radiotherapy in various stages of treatment
Drug therapy	Drugs and chemicals, including chemotherapy, targeted therapy, and endocrine therapy	The new drug pertuzumab (Perjeta) has been used together with herceptin and chemotherapy to shrink tumors completely, so some patients do not need surgery
Other therapies	Other therapies, including traditional Chinese medicine, biotherapy, and interventional therapy	With the application of traditional Chinese medicine, most patients with breast cancer will not need surgery

[a] HPV: human papillomavirus.

Diffusion Characteristics of Gynecologic Cancer–Related Information

Among all of the 524 tweets categorized as misinformation, only 64 received retweet or retweets (mean 5.09 [SD 70.40]), 72 received comment or comments (mean 0.53 [SD 2.62]), and 132 received like or likes (mean 2.68 [SD 31.80]). The popularity of misinformation in terms of its diffusion was unevenly distributed, with several tweets receiving a large number of retweets, whereas the majority received no retweets or likes. For instance, the 2 most popular false tweets were about gynecologic cancer prevention and treatment methods that involved eating specific foods, such as garlic, mushrooms, and red wine; these received 1143 and 1131 retweets, respectively.

Among the 1023 tweets categorized as true information, only 143 received retweet or retweets (mean 4.06 [SD 32.58]). A total of 120 tweets received comment or comments (mean 1.78 [SD 14.41]), and 167 tweets received like or likes (mean 2.54 [SD 28.08]). Similarly, most of the tweets vanished into obscurity after being published, and several tweets reached a high degree of popularity. Moreover, most of these popular tweets were about cancer prevention methods, such as lifestyle and vaccinations. Figure 1 displays the retweet network of all the cancer-related true information and misinformation as well as the largest retweet networks for both true information and misinformation.

By comparing the diffusion characteristics of the true information and misinformation, it was found that true information was generally better diffused and accepted than misinformation by social media users. Figure 2 shows the complementary cumulative distribution functions of the 5 diffusion characteristics of true information and misinformation. Although several false tweets had been extremely popular and received a large number of retweets, most of the false tweets received less retweets than the true tweets. In addition, true tweets had better diffusion performance in terms of the range and structural virality of retweet networks and the number of comments. All of the indices showed that true information spread more deeply and broadly than misinformation, reaching a larger audience on social media.

A between-subject multivariate analysis of variance (MANOVA) was performed to test the differences on the 5 diffusion characteristics between information in different thematic categories and those with different information veracity. Information in the diagnosis category was excluded in the analysis, as there were not adequate cases in the misinformation group (n=14).

The MANOVA results show that the interaction effect between thematic category and information veracity was significant ($F_{2,1399}$=3.26, P<.001; Wilks lambda=0.98;η p^2=.011). Figure 3 reports the estimated means of 5 diffusion characteristics adjusted by information veracity and thematic category, as well as their CI at 95% confidence level.

Specifically, regarding the thematic category of background knowledge, there were significant differences between the true information and misinformation groups in the scale, range, and structural virality of retweet networks, although there were nonsignificant differences in the number of likes and number of comments. It showed that individuals were more likely to spread true information about background knowledge than misinformation.

In terms of the thematic category of treatment, although estimated means revealed that true information generally had higher diffusion indices than misinformation, these differences were not significant. In addition, for prevention-related information, significant differences between the true information and misinformation groups occurred in the scale and range of retweets as well as the number of likes. There was no significant difference in the number of comments and structural virality of retweets. According to Figure 3, prevention-related misinformation spread better than true information on social media.

Figure 1. An illustration of the retweet network: (A) the full retweet network of all true information (red) and misinformation (green); (B) the largest retweet network of true information; (C) the largest retweet network of misinformation.

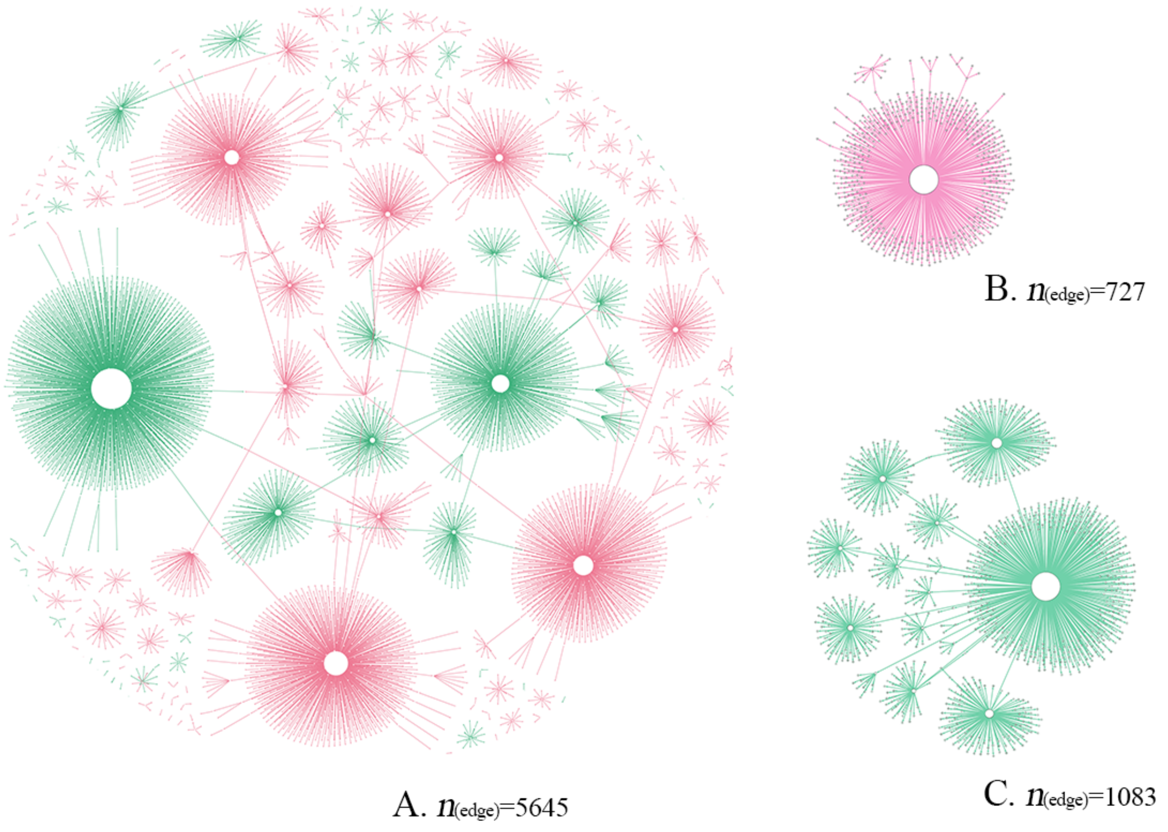


Figure 2. Complementary cumulative distribution functions of true information and misinformation cascades (x-axis and y-axis are log-transformed).

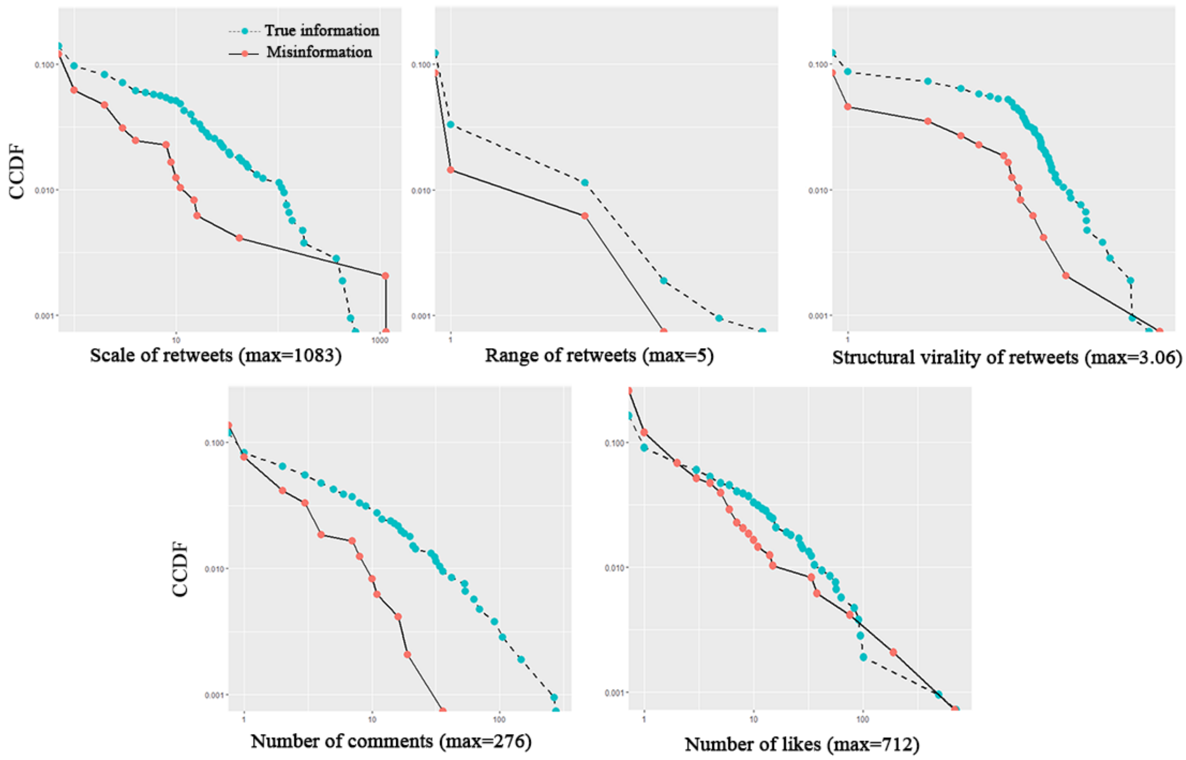
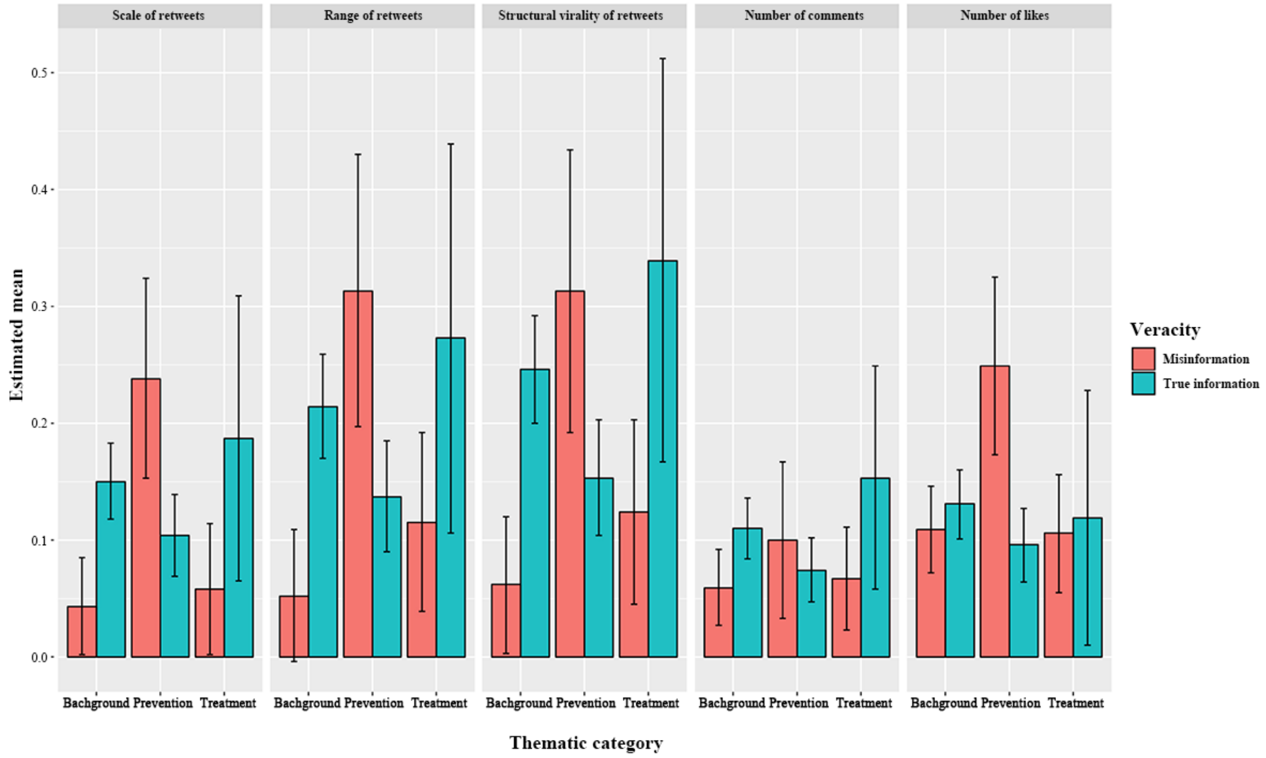


Figure 3. Estimated diffusion characteristics adjusted by thematic category and information veracity. Means reported here are estimated marginal means of multivariate analysis of variance (MANOVA). Outliers, multivariate normality, linear relationships between dependent variables, and multicollinearity were checked before analysis. Scale of retweets, number of comments, number of likes were log-transformed to fit the assumption of normal distribution.



Discussion

Principal Findings

Individuals have increasingly used social media to exchange cancer-related information[5,26]. However, such information may includemany false elements, and these could distort individuals' attitudes and behaviors toward cancer prevention and treatment. This study seeks to understand thenature and diffusion of cancer-related misinformation on social media.

First, the findings revealed that of the 2691 total tweets examined, more than half included medically oriented information about cancer. Although most of the medically oriented tweets provided accurate information, more than 30% contained misinformation. This finding suggests that the public should assess the information veracity of tweets on social media before accepting and following the advice embedded in such information. Moreover, the large amount of cancer-related misinformation on social media suggests the need for correction and reduction of misinformation, with the efforts from both social media service providers and medical professionals. For instance, with the knowledge from medical professionals, social media service providers could establish a digital library for cancer-related misinformation, which could provide a scaffold for self-checking by the public [37].

In addition, the results indicated that social media tweets related to cancer treatment contained a substantially greater percentage of misinformation than true information. However, the network analysis of the information diffusions showed that cancer treatment–related misinformation did not outperform the true information in terms of their diffusion characteristics. Moreover, for background knowledge, true information spread to a wider range of audiences than misinformation. These findings were inconsistent with those of Vosoughi et al's study [19] in which misinformation diffused better than true information. The inconsistency may be caused by the unique context of gynecologic cancers. Unlike Vosoughi et al [19], who examined all types of news stories on Twitter, we focused on misinformation related to 2 specific types of gynecologic cancers only. Due to the wide media coverage of female celebrities diagnosed with breast and cervical cancers, many Chinese people are well aware of the topic [38], thus limiting the spread of misinformation related to it on social media. This implies that future research on the spread of misinformation on Web should adopt a topic-specific or domain-specific approach.

More interestingly, the diffusion characteristics of prevention-related misinformation are quite different from the information in other thematic categories. Although there was a relatively small amount of prevention-related misinformation on social media, this misinformation diffused significantly more broadly and deeply than true information. One possible explanation could be that a large amount of prevention-related misinformation provided ways or actions to prevent breast cancer and cervical cancer that individuals could perform by themselves. In other words, the prevention-related misinformation contained both self-efficacy and response efficacy, which could help individuals reduce anxiety and fear as well as control their perceived threat from cancer [39]. Thus, individuals are more willing to spread these prevention-related messages.

Implications

Several practical implications can be derived here. First, medical professionals should make efforts to correct misinformation regarding the appropriate ways of preventing gynecologic cancers and decrease the spread of cancer-preventing misinformation on social media. This calls for the establishment of online health communities to list common cancer-related misinformation and provide accurate information about cancer prevention to address the public misperceptions of cancer [40]. Second, the government should run health campaigns and education programs to improve the public's health literacy and strengthen their capacities to obtain, read, understand, and assess health care information so that they can use Web-based health information effectively and make appropriate health decisions [41,42]. Third, the public should be encouraged to verify the accuracy of Web-based cancer-related information, especially preventive information such as superfoods and vaccinations, to protect them from using counterfeit, inappropriate, or unsafe cancer prevention measures, as suggested by Bode and Vraga (2018) [43]. Finally, the significant difference between the diffusion characteristics of true information and misinformation implies that those diffusion characteristics can act as heuristics to identify cancer misinformation. In other words, the spread of misinformation on Web usually follows specific patterns that are different from true information, suggesting that interventions can come into play in the early stage of misinformation diffusion [19].

Limitations

Several limitations of this study should be acknowledged. First, this study only focused on gynecologic cancer–related information. The public's awareness and knowledge about specific diseases and health issues may vary to a different extent, which might affect how they perceive and disseminate relevant misinformation on social media. Thus, future studies could examine the nature and diffusion of misinformation regarding other disease or health issues in different cultural settings.

Second, although this study provided an empirical investigation on the types of gynecologic cancer–related misinformation and its diffusion characteristics, the factors driving the diffusion of different types of misinformation remain unknown. Future research should examine the mechanisms behind the diffusion of misinformation on social media and elucidate effective strategies for curbing the spread of misinformation. Finally, the data were extracted in July 2016, which were 2 years old. With the increasing popularity of social media, people's health literacy has been continuously improving over the last few years. They are increasingly cautious about posting and sharing health information on social media, which might change the nature and diffusion of cancer information. Thus, future research with a wider time span could be conducted to investigate the inherent changes of cancer information diffusion on social media.

Conclusions

This study makes the first attempt to examine the nature and diffusion of cancer-related misinformation on Chinese social media. First, the gynecologic cancer–related tweets were content-analyzed to differentiate between true information and misinformation on social media as well as to identify the types of falsehoods. In addition, a network perspective was adopted to examine the diffusion characteristics of misinformation through comparisons with those of true information. The results indicated that although most of the gynecologic cancer–related tweets provided medically accurate information, approximately 30% contained misinformation.

More importantly, although cancer treatment–related tweets included a great amount of misinformation, the misinformation did not diffuse significantly greater than true information. Conversely, cancer prevention tweets contained a relatively small amount of misinformation, but it spread more broadly and deeply than true information. These findings suggest that the government, social media service providers, and medical professionals should make great efforts to decrease the prevalence of cancer misinformation on social media. Moreover, health campaigns and programs should be conducted to increase the public's motivations and abilities to verify Web-based cancer-related information, especially preventive measures before sharing or following the instructions from these messages.

Conflicts of Interest

None declared.

References

1. Chen WQ, Li H, Sun KX, Zheng RS, Zhang SW, Zeng HM, et al. [Report of cancer incidence and mortality in China, 2014]. Zhonghua Zhong Liu Za Zhi 2018 Jan 23;40(1):5-13. [Medline: 29365411]
2. Jiang X, Tang H, Chen T. Epidemiology of gynecologic cancers in China. J Gynecol Oncol 2018 Jan;29(1):e7 [FREE Full text] [doi: 10.3802/jgo.2018.29.e7] [Medline: 29185265]
3. Shi J, Chen L. Social support on Weibo for people living with HIV/AIDS in China: a quantitative content analysis. Chin J Commun 2014 Jul 24;7(3):285-298. [doi: 10.1080/17544750.2014.926954]
4. Chen L, Yang X. Nature and effectiveness of online social support for intercultural adaptation of mainland Chinese international students. Int J Commun 2015;9:2161-2181.
5. Himelboim I, Han JY. Cancer talk on twitter: community structure and information sources in breast and prostate cancer social networks. J Health Commun 2014;19(2):210-225. [doi: 10.1080/10810730.2013.811321] [Medline: 24111482]
6. Yan YY. Online health information seeking behavior in Hong Kong: an exploratory study. J Med Syst 2010 Apr;34(2):147-153. [Medline: 20433053]
7. Song H, Omori K, Kim J, Tenzek KE, Morey HJ, Lin WY, et al. Trusting social media as a source of health information: online surveys comparing the United States, Korea, and Hong Kong. J Med Internet Res 2016 Mar 14;18(3):e25 [FREE Full text] [doi: 10.2196/jmir.4193] [Medline: 26976273]
8. 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 Mar;28(2):361-368 [FREE Full text] [doi: 10.1377/hlthaff.28.2.361] [Medline: 19275991]
9. Ayers SL, Kronenfeld JJ. Chronic illness and health-seeking information on the internet. Health (London) 2007 Jul;11(3):327-347. [doi: 10.1177/1363459307077547] [Medline: 17606698]
10. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. Br Med J 2002 Mar 9;324(7337):573-577 [FREE Full text] [Medline: 11884321]
11. Syed-Abdul S, Fernandez-Luque L, Jian WS, Li YC, Crain S, Hsu MH, et al. Misleading health-related information promoted through video-based social media: anorexia on YouTube. J Med Internet Res 2013 Feb 13;15(2):e30 [FREE Full text] [doi: 10.2196/jmir.2237] [Medline: 23406655]
12. Vraga EK, Bode L. Using expert sources to correct health misinformation in social media. Sci Commun 2017 Sep 14;39(5):621-645. [doi: 10.1177/1075547017731776]
13. Vogel L. Viral misinformation threatens public health. CMAJ 2017 Dec 18;189(50):E1567 [FREE Full text] [doi: 10.1503/cmaj.109-5536] [Medline: 29255104]
14. Søe SO. Algorithmic detection of misinformation and disinformation: Gricean perspectives. J Doc 2018 Mar 12;74(2):309-332. [doi: 10.1108/JD-05-2017-0075]
15. Allcott H, Gentzkow M. Social media and fake news in the 2016 election. J Econ Perspect 2017;31(2):211-236 [FREE Full text]
16. Anderson J, Rainie L, Caiazza T. Elon.edu. The future of truth and misinformation online URL: http://www.elon.edu/docs/e-web/imagining/surveys/2017_survey/Future_of_Info_Environment_Elon_University_Pew_10-18-17.pdf [accessed 2018-10-07] [WebCite Cache ID 72zLj5d3o]
17. Fernández-Luque L, Bau T. Health and social media: perfect storm of information. Healthc Inform Res 2015 Apr;21(2):67-73 [FREE Full text] [doi: 10.4258/hir.2015.21.2.67] [Medline: 25995958]
18. Del Vicario M, Bessi A, Zollo F, Petroni F, Scala A, Caldarelli G, et al. The spreading of misinformation online. Proc Natl Acad Sci U S A 2016 Jan 19;113(3):554-559 [FREE Full text] [doi: 10.1073/pnas.1517441113] [Medline: 26729863]

19. Vosoughi S, Roy D, Aral S. The spread of true and false news online. Science 2018 Dec 09;359(6380):1146-1151. [doi: 10.1126/science.aap9559] [Medline: 29590045]
20. Lewandowsky S, Ecker UK, Seifert CM, Schwarz N, Cook J. Misinformation and its correction: continued influence and successful debiasing. Psychol Sci Public Interest 2012 Dec;13(3):106-131. [doi: 10.1177/1529100612451018] [Medline: 26173286]
21. Thai M, Wu W, Xiong H. Big Data in Complex and Social Networks. London: Chapman and Hall/CRC Press; 2016.
22. Oh O, Agrawal M, Rao HR. Community intelligence and social media services: a rumor theoretic analysis of tweets during social crises. MIS Q 2013 Feb 2;37(2):407-426. [doi: 10.25300/MISQ/2013/37.2.05]
23. De Domenic M, Lima A, Mougel P, Musolesi M. The anatomy of a scientific rumor. Sci Rep 2013 Oct 18;3:2980 [FREE Full text] [doi: 10.1038/srep02980] [Medline: 24135961]
24. Zhao L, Yin J, Song Y. An exploration of rumor combating behavior on social media in the context of social crises. Comput Human Behav 2016 May;58:25-36. [doi: 10.1016/j.chb.2015.11.054]
25. Russell C. Living can be hazardous to your health: how the news media cover cancer risks. J Natl Cancer Inst Monogr 1999 Jan 01;1999(25):167-170. [doi: 10.1093/oxfordjournals.jncimonographs.a024194]
26. Gage-Bouchard EA, LaValley S, Warunek M, Beaupin LK, Mollica M. Is cancer information exchanged on social media scientifically accurate? J Cancer Educ 2017 Jul 19. [doi: 10.1007/s13187-017-1254-z] [Medline: 28721645]
27. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. CA Cancer J Clin 2016;66(2):115-132 [FREE Full text] [doi: 10.3322/caac.21338] [Medline: 26808342]
28. World Health Organization. 2002. National Cancer Control Programmes: Policies and Managerial Guidelines URL: http://www.who.int/cancer/media/en/408.pdf [accessed 2018-10-07] [WebCite Cache ID 72zNzP3Zu]
29. Bidmon S, Terlutter R. Gender differences in searching for health information on the Internet and the virtual patient-physician relationship in Germany: exploratory results on how men and women differ and why. J Med Internet Res 2015;17(6):e156 [FREE Full text] [doi: 10.2196/jmir.4127] [Medline: 26099325]
30. Krippendorff K. Agreement and information in the reliability of coding. Commun Methods Meas 2011 Apr;5(2):93-112. [doi: 10.1080/19312458.2011.568376]
31. 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]
32. Tesch R. Qualitative Research: Analysis Types and Software Tools. New York: Falmer Press; 1990.
33. Morse JM, Field PA. Qualitative Research Methods for Health Professionals. New York: Sage Publications; 1995.
34. Bakhshi B, Shamma DA, Gilbert E. Faces Engage Us Photos with Faces Attract More Likes and Comments on Instagram. 2014 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 26-May 01, 2014; Toronto, Ontario, Canada p. 965-974. [doi: 10.1145/2556288.2557403]
35. Yang J, Counts S. Predicting the speed, scale, and range of information diffusion in Twitter. 2010 Presented at: Fourth International Conference on Weblogs and Social Media 2010; May 23-26, 2010; Washington, DC, US p. 355-358.
36. Goel S, Anderson A, Hofman J, Watts DJ. The structural virality of online diffusion. Manage Sci 2015 Jul 22;62(1):180-196. [doi: 10.1287/mnsc.2015.2158]
37. Lazer DM, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, et al. The science of fake news. Science 2018 Dec 09;359(6380):1094-1096. [doi: 10.1126/science.aao2998] [Medline: 29590025]
38. Leung VS, Cheng K. Public's perception of celebrities with serious illness in Hong Kong and the impact of media stories of ill celebrities on health awareness and behaviour. J Commun Healthc 2016 Oct 26;9(4):256-266. [doi: 10.1080/17538068.2016.1247128]
39. Tannenbaum MB, Hepler J, Zimmerman RS, Saul L, Jacobs S, Wilson K, et al. Appealing to fear: a meta-analysis of fear appeal effectiveness and theories. Psychol Bull 2015 Nov;141(6):1178-1204. [doi: 10.1037/a0039729] [Medline: 26501228]
40. Kanthawala S, Vermeesch A, Given B, Huh J. Answers to health questions: Internet search results versus online health community responses. J Med Internet Res 2016 Apr 28;18(4):e95 [FREE Full text] [doi: 10.2196/jmir.5369] [Medline: 27125622]
41. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health literacy and health information technology adoption: the potential for a new digital divide. J Med Internet Res 2016 Oct 04;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]
42. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, (HLS-EU) Consortium Health Literacy Project European. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012 Jan 25;12:80 [FREE Full text] [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]
43. Bode L, Vraga EK. See something, say something: correction of global health misinformation on social media. Health Commun 2018 Sep;33(9):1131-1140. [doi: 10.1080/10410236.2017.1331312] [Medline: 28622038]

Abbreviations

MANOVA: multivariate analysis of variance

Edited by G Eysenbach; submitted 07.07.18; peer-reviewed by W Pian, S Syed Abdul, L Fernandez-Luque; comments to author 06.08.18; revised version received 02.09.18; accepted 01.10.18; published 16.10.18.

Please cite as:
Chen L, Wang X, Peng TQ
Nature and Diffusion of Gynecologic Cancer–Related Misinformation on Social Media: Analysis of Tweets
J Med Internet Res 2018;20(10):e11515
URL: http://www.jmir.org/2018/10/e11515/
doi:10.2196/11515
PMID:30327289

Original Paper

Using Twitter to Examine Web-Based Patient Experience Sentiments in the United States: Longitudinal Study

Kara C Sewalk[1*], MPH; Gaurav Tuli[1*], PhD; Yulin Hswen[1,2], MPH; John S Brownstein[1,3,4], PhD; Jared B Hawkins[1,3], MMSc, PhD

[1]Computational Health Informatics Program, Boston Children's Hospital, Boston, MA, United States
[2]Department of Social and Behavioral Sciences, Harvard TH Chan School of Public Health, Boston, MA, United States
[3]Department of Pediatrics, Harvard Medical School, Boston, MA, United States
[4]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, United States
[*]these authors contributed equally

Corresponding Author:
Kara C Sewalk, MPH
Computational Health Informatics Program
Boston Children's Hospital
BCH3393
300 Longwood Avenue
Boston, MA, 02115
United States
Phone: 1 8572185188
Email: kara.sewalk@childrens.harvard.edu

Abstract

Background: There are documented differences in access to health care across the United States. Previous research indicates that Web-based data regarding patient experiences and opinions of health care are available from Twitter. Sentiment analyses of Twitter data can be used to examine differences in patient views of health care across the United States.

Objective: The objective of our study was to provide a characterization of patient experience sentiments across the United States on Twitter over a 4-year period.

Methods: Using data from Twitter, we developed a set of 4 software components to automatically label and examine a database of tweets discussing patient experience. The set includes a classifier to determine patient experience tweets, a geolocation inference engine for social data, a modified sentiment classifier, and an engine to determine if the tweet is from a metropolitan or nonmetropolitan area in the United States. Using the information retrieved, we conducted spatial and temporal examinations of tweet sentiments at national and regional levels. We examined trends in the time of the day and that of the week when tweets were posted. Statistical analyses were conducted to determine if any differences existed between the discussions of patient experience in metropolitan and nonmetropolitan areas.

Results: We collected 27.3 million tweets between February 1, 2013 and February 28, 2017, using a set of patient experience-related keywords; the classifier was able to identify 2,759,257 tweets labeled as patient experience. We identified the approximate location of 31.76% (876,384/2,759,257) patient experience tweets using a geolocation classifier to conduct spatial analyses. At the national level, we observed 27.83% (243,903/876,384) positive patient experience tweets, 36.22% (317,445/876,384) neutral patient experience tweets, and 35.95% (315,036/876,384) negative patient experience tweets. There were slight differences in tweet sentiments across all regions of the United States during the 4-year study period. We found the average sentiment polarity shifted toward less negative over the study period across all the regions of the United States. We observed the sentiment of tweets to have a lower negative fraction during daytime hours, whereas the sentiment of tweets posted between 8 pm and 10 am had a higher negative fraction. Nationally, sentiment scores for tweets in metropolitan areas were found to be more extremely negative and mildly positive compared with tweets in nonmetropolitan areas. This result is statistically significant ($P<.001$). Tweets with extremely negative sentiments had a medium effect size ($d=0.34$) at the national level.

Conclusions: This study presents methodologies for a deeper understanding of Web-based discussion related to patient experience across space and time and demonstrates how Twitter can provide a unique and unsolicited perspective from users on the health care they receive in the United States.

(J Med Internet Res 2018;20(10):e10043) doi:10.2196/10043

KEYWORDS

health care; social media; patient experience

Introduction

In the past decade, we have observed a shift in the United States health care system to emphasize a patient-centered approach to care [1]. Standardized practices to qualitatively assess the care patients receive at hospitals have been developed, such as the Hospital Consumer Assessment of Healthcare Providers and Systems survey [2]. Many benefits to patient-centered health care facilities have been identified, including reduced length of stay, lower costs per case, decreased adverse events, and even reduced operating costs [1]. Studies have even found that better reported patient care experiences are associated with better clinical outcomes, improved safety within hospitals, and less frequent use of health care [3,4].

Traditional assessments have also documented differences in access to health care [5]. Research has shown that access to health care varies based on where a patient lives [6,7,8]. Patient care is often dependent upon the policies of the state a patient lives in, distance to the nearest health care facilities, and insurance coverage, which varies across the United States. Population size can impact many of these factors. It has been shown that individuals in large metropolitan cities tend to have better access and quality of care compared with smaller, more rural communities [6].

However, commonly used assessments of patient care, such as surveys or focus groups, have limitations that include social desirability bias, smaller audiences, and restrictions on what questions and topics patients are asked about [9,10]. The Pew Research Center reported that 87% of Americans who have seen a health care provider report positive feedback on their experience. However, 39% of US adults believe that US health care is below average [11,12].

With an increasing demand for transparency in health care, social media has shifted to become a platform for patient engagement and empowerment. Currently, there are 69 million monthly active Twitter users in the United States [13], highlighting the overwhelming use and potential for rich information to be extracted from the social networking site. Information on social media could be valuable to complement evaluations of patient care because Web-based posts provide an unsolicited, free-text perspective from users on the care they receive. There are limited studies which provide in-depth examinations of care across the United States and few, if any, that are reflections of social media discussions.

Previous research has shown that Twitter can be used as a supplemental data stream for measuring the patient-perceived quality of care in US hospitals by comparing patient sentiments about hospitals with established quality measures and traditional hospital-based feedback reports [14]. This indicates that Web-based data about patient experience and hospital care that is valuable to explore further are available from Twitter. Additionally, such research has shown that a range of topics can be identified and understood from these tweets [14,15]. Novel approaches can be used to further describe differences in hospital performances [16]. This includes sentiment analysis, a process that examines the content of free-short message service text messages and determines a score rating on a scale of positive to negative [17]. Sentiment analyses have been shown useful in describing patient opinions on hospital care that are comparable with results from more traditional survey methods [18]. An evaluation of research using sentiment analyses for health care-related tweets identified a need for improved methods of understanding sentiment data in a health care setting [17]. Previous examinations have also shown that social media research has explored specific public health topics and target populations, but there lacks a comprehensive study that fully examines a communication tool for a larger scope to evaluate population health needs [19].

To examine sentiments of health care in the United States online, we captured tweets discussing patient experiences not restricted by the level or type of health service provided. This dataset is the first of its kind that explores carefully curated data from the Twitter platform related to patient experience, which includes, but is not limited to, interactions at hospitals, urgent care facilities, primary and specialty care offices, and related health care facilities. Using this rich dataset, we aim to provide a spatial and temporal characterization of the sentiment of health care discussions on Twitter and determine if there are differences in the sentiment of health care discussions between metropolitan and nonmetropolitan areas in the United States using Twitter as a real-time supplementary data stream. Insight on patient experience discussions online can help inform health care facilities, key stakeholders and future research practices for examining patient feedback using Web-based data.

Methods

Patient Experience Classifier

This study utilizes data from the social media platform Twitter to investigate the experiences of patients at hospitals, urgent care facilities, primary and specialty care offices, and other related health care facilities. We used a combination of keywords to gather publicly available patient experience-related tweets through Gnip, the Twitter-owned data broker. Gnip is a paid licensing software service for Twitter data. All data collected in this study were publicly posted on Twitter; therefore, per the privacy policy of Twitter [20], users elect to have this information available to the general public for consumption. A set of keywords and rules were meticulously chosen to retrieve tweets potentially discussing experience related to the following areas: medical facilities and staff, medical procedures, hospital visits and stays, medications, hospital bills and insurance, care condition, and pain. The keywords were divided into the classes to correctly form the rules. A list of classes along with the corresponding set of keywords and example rules are shown in Multimedia Appendix

1; for example, care condition keywords include monitor, heal, recover, care, cure, dying, dead, sicker, sick, ill, illness, and condition. The keywords retrieved 27,309,724 unique tweets (45.3 million when including the retweets) posted between February 1, 2013 and February 28, 2017. The retweets were not considered in the study.

We developed a set of software components to auto label and examine the patient experience Twitter dataset. The set includes a classifier to determine patient experience tweets, a geolocation inference engine for social data, a modified version of a sentiment classifier from the literature, and an engine to determine if the tweet is from a metropolitan or nonmetropolitan area. These components were built for appropriately handling health care experience social data.

For the purpose of this study, we identified the tweets captured that were relevant to the patient experience. A relevant tweet included discussions about care received in a hospital, urgent care, or any other health institution—either by the person themselves, a friend, or relative. We aimed to capture tweets that discussed any exposure to health care.

We built a supervised machine classifier for identifying relevant patient experience tweets. A 2-step curation approach was adopted to create a training dataset for the classifier. We determined that tweets containing a Web page link (also known as URL) are 18 times more likely to be irrelevant. Two randomly selected sets of 5000 tweets, one with and the other without URLs, were hand curated using Amazon Mechanical Turk (MTurk) for this examination. The set with URLs contained only 56 of 4599 agreed upon relevant tweets (1.22%) compared with 760 of 3439 agreed upon relevant tweets (22.10%) in the set without URLs. Therefore, we decided to only consider tweets without URLs for this study. We curated 15,000 additional tweets without URLs on MTurk. In total, the manual MTurk curation gave us 3708 relevant and 9810 irrelevant patient experience tweets for which at least two of the MTurk curators were in agreement. There was an agreement on a total of 13,885 of 20,000 tweets without URLs (69.43%) between the MTurk curators. All MTurk curators selected were identified as master's-level workers, having been monitored and verified by Amazon as high performing and demonstrating excellence in their curation tasks [21]. All MTurk curators were restricted to only curate each tweet once. Example curation instructions for the MTurk curators are presented in Multimedia Appendix 2. A few examples of manually curated tweets are shown in Table 1. The tweets provided in this table are fictitious examples to preserve user identity and privacy, a technique that has been recommended in previous research to address the ethical concerns of disseminating Twitter data [22].

We developed a support vector machine-based supervised machine learning classifier using this training set to filter relevant tweets from irrelevant ones. The classifier was built using various textual features and was iteratively evaluated using the 10-fold cross validation over 90% training and 10% test sets. Each training tweet was tokenized using the Natural Language Toolkit TweetTokenizer. Stop words and mentions (ie, words beginning with "@'") were removed. Unigrams and bigrams with term frequency-inverse document frequency normalization were used as features. Other features included whether the tweet contained a reference to a hospital staff member and a reference to themselves or a family member. We selected the top 15,000 features from a classifier that produced the highest F_1 score with the lowest overfitting. The classifier was assessed for overfitting by comparing the difference in the performance on the training and test sets.

Geolocation of Tweets

This study aims to analyze and compare patient experience sentiments at national and regional levels in the country using the Twitter data. However, Twitter data very rarely contain location information. Previous studies have found that a very small fraction of users share their geo-coordinates in the tweets [23]. We also found that only 2.97% (81,930/2,759,257) of the total relevant patient experience tweets contained user-defined geo-coordinates. Therefore, we developed a location inference engine to approximately identify geographical locations, such as country, state, and region of the relevant tweets in this dataset.

We used a combination of information from the users' profile and GPS (Global Positioning System) coordinates of tweets, when available, to infer the location of the tweets. We also used the Google Maps Geocoding application programming interface [24] in conjunction with the US Census Bureau state boundaries [25] to infer the US state of each tweet. Because a user can input any free text containing a combination of words, symbols, and emojis as location in their profile, we built a library of highly used junk locations (eg, "in your heart," "with aliens," "under your bed," etc) combined with natural language processing (NLP) to identify useless location strings. A list of example location strings is shown in Multimedia Appendix 3. Our geolocation engine was built specifically for social media, wherein users are free to provide any string as their location. It augments Google's geocoding service [26] with NLP and data mining. This engine performs a list of NLP operations to get rid of irrelevant locations and to parse and format location strings followed by querying to Google Map application programming interface for geolocating the location. We chose to use Google's geocoding service because it has been repeatedly reported to have a better accuracy [27], thorough coverage [28], and is equipped to handle ambiguous locations [26].

Table 1. Example tweets for the patient experience dataset curation.

Tweet class	Examples
Patient experience[a]; hand labeled: 3708	After having a tumor removed from my bladder I returned to the ward with a catheter fitted. #cityhospital
Irrelevant[b]; hand labeled: 9810	Need tips for better communication with your doctor? #medicine #wellness

[a] Patient experience tweets are defined as tweets related to an individual's experience in a health care setting.

[b] Irrelevant tweets are any tweets captured in the database that are not patient experience tweets.

There are other geocoding services, such as Nominatim and Carmen, which could have been used in this study. However, there is a limitation to using Nominatim because geolocation is tightly coupled to specific address formats [29], which would be difficult to use with Twitter data because users can specify their location in any format using a free-text field. Additionally, Carmen provides maximum resolution only at the city level for both geocoding and reverse geocoding, which may lead to incorrect results for the users who provide finer-grained locations such as neighborhoods [30]. The location database and alias list of Carmen also needs improvement. The creators of Carmen recommend augmenting the location database and alias list by querying to other search engines and public resources [30]. For this reason, we found that the geolocation engine we built is better suited for the purposes of this study.

Using the geolocation engine, we determined the state location for each tweet and the associated broader region that each state was assigned to. The regions examined in this study were chosen and aggregated as defined by the US Census Bureau, which are each a grouping of states and identified with a single-digit census code [31]. The US Census Bureau groups each region by similarities in historical development, population, and economy and recommends using this framework for comparative efforts [32]. Previous research has shown regional differences in health care [33], and this study sought to determine if regional differences in care could be identified on Twitter. Further details of the tweet extraction, curation, and geocoding are provided in Multimedia Appendix 4.

Tweet Sentiment

A prime objective of this study is to gauge and compare the sentiments of patient experiences across the country. To compute the sentiments expressed in the tweets, we adopted a widely accepted and used lexicon and rule-based sentiment classifier called Valence Aware Dictionary for Sentiment Reasoning (VADER) [34]. However, we appended VADER's dictionary and rules to provide a broader representation of Twitter data, which included incorporating more than 110 emojis and their respective sentiment scores [35].

VADER computes sentiment and valence for each word level and provides positive, negative, and neutral scores at the sentence level. We used the compound score, which is a unidimensional and normalized measure of sentiment. It is computed by summing the valence scores of each word in the lexicon, adjusted according to the rules, and normalized to be between −1 (most extreme negative) and +1 (most extreme positive). We used VADER to compute the compound sentiment score for every sentence in the tweet, and then took the mean of all nonzero compound scores to provide a sentiment score per tweet. We considered a sentiment positive if the mean compound score was ≥0.3 or negative if the score was ≤−0.3. Mean compound scores between −0.3 and 0.3 were considered neutral. In the majority of our analyses for this study, we considered tweets with positive and negative scores only because these sentiments provide more actionable data.

Population Size Examination

We further explored if the patterns of discussion and reporting about patient experience vary by geographical region and by population size of the location of the Twitter users. To perform this analysis, we aggregated the labeled Twitter data with identified state locations into 4 US regions and also dichotomized the data into metropolitan (population ≥50,000) and nonmetropolitan (population <50,000) areas [6].

We used the recent and most detailed geographic polygon data on urban areas from the US Census Bureau [36] to infer if a tweet was from a metropolitan or nonmetropolitan area. According to these data, there are more than 486 urbanized areas (population ≥50,000) and 3087 urban clusters (5,000 ≤population<50,000) in the United States, accounting for a total of 24,356 geographic polygons. The geo-coordinate of each tweet inferred by our location identification engine was checked against these polygons. A tweet was considered metropolitan if the geo-coordinate of the tweet fell inside a geographic polygon of an urbanized area. The tweets falling either inside a polygon of the urban clusters or falling outside all of the urban polygons were considered nonmetropolitan tweets.

Temporal Examination

The time at which a tweet is posted can be an informative dimension to analyze the patient experience. Certain sentiment patterns, for example, might be more popular during the day than at night. To uncover such patterns, we analyzed Twitter data regarding patient experience by examining the time of the day and that of the week when the tweets were posted. This gives us a broad set of trends to analyze the activity of a selected geographic region.

Because the timestamps of Twitter data are provided in coordinated universal time (also known as Greenwich Mean Time), this analysis requires converting the time at which a tweet was posted onto a Twitter user's local time. We used the inferred state information provided by sour geolocation classifier along with the time zone information for each state to identify the correct coordinated universal time offset to calculate the local time.

Statistical Analysis

To determine if there were any differences between the discussions of patient experience on Twitter in metropolitan and nonmetropolitan areas, we performed a Mann-Whitney nonparametric test on the sentiment scores of the tweets. We tested the ranked distribution of metropolitan and nonmetropolitan sentiment scores to determine if they were approximately equal at national and regional levels, aggregating positive and negative scores together. We also compared the metropolitan and nonmetropolitan sentiment scores at national and regional levels by the sentiment polarity and valence. The nonparametric tests were chosen because the sentiment score distribution was found to be symmetric and bimodal.

Results

Geolocation of Tweets

After evaluating a set of classifiers, we selected a support vector machine classifier that produced the highest F_1 score with the lowest overfitting. The selected classifier achieved an accuracy of 83% with a precision and recall of 70% and 69%, respectively, for the patient experience tweet class. We filtered the gathered tweets with no URLs and ran the selected classifier to identify patient experience tweets. There were 33.88% of the total tweets (9,252,004/27,309,724) found to be without a URL, out of which 29.82% (2,759,257/9,252,004) were labeled as patient experience by the classifier. We also verified the classifier-labeled patient experience tweets by manually curating a random set of 5000 tweets and found it to be 76% in agreement with the classifier.

To perform national and regional analyses, the labeled patient experience tweets were required to be geocoded. We found that only 2.97% (81,930/2,759,257) of the total patient experience tweets contained geo-coordinates shared by the users. After using our geolocation inference engine, we identified 31.76% (876,384/2,759,257) patient experience tweets that belonged to 1 of the 50 US states, District of Columbia, Puerto Rico, or the United States Virgin Islands; 19.25% (531,062/2,759,257) of the patient experience tweets were from outside the United States, whereas 14.58% (402,295/2,759,257) had insufficient information and 35.14% (969,614/2,759,257) had no information to infer geolocation. Manual curation of 10,000 randomly selected tweets using MTurk validated that 91% (9100/10,000) of the inferred locations through the geolocation engine were correct (with 87%, 8,700/10,000 agreement between 2 MTurk curators). We also verified the quality of the MTurk curators for this task using an in-house team to manually curate the first 2000 tweet geolocations. Our curators had 79% agreement with the MTurk curators.

The further dichotomization of the patient experience Twitter dataset into metropolitan and nonmetropolitan tweets identified 69.36% (607,891/876,384) of total tweets as metropolitan tweets and 30.64% (268,493/876,384) as nonmetropolitan tweets across the 4-year study period from February 2013 to February 2017. The state of Rhode Island was identified as the state with most tweets in a metropolitan area per 100,000 residents (at 97.2%) in the patient experience dataset, and Wyoming had the most tweets in a nonmetropolitan area per 100,000 residents in the patient experience dataset (at 89.7%); 100% of the tweets from the District of Columbia were metropolitan because it is entirely urbanized.

Tweet Sentiment

Of the 27,309,724 tweets collected between February 2013 and February 2017 using a set of patient experience-related keywords, the classifier was able to identify 2,759,257 tweets that were labeled as patient experience. After running the patient experience tweets through the geolocation classifier, we identified 876,384 tweets by approximate location to use for spatial analyses. At the national level, we observed 27.83% positive (243,903/876,384), 36.22% neutral (317,445/876,384), and 35.95% negative (315,036/876,384) patient experience tweets in the dataset. For this study, we chose to exclude tweets with neutral sentiment scores.

Figure 1 and Figure 2 show the patient experience tweet count and sentiment trends over the 4-year study period across the 4 regions of the United States. The color scale of the 4 regions in Figure 1 represents the average sentiment polarity rate and the blue dot in each state depicts the approximate size of the patient experience tweet rate.

The average sentiment polarity rate is the mean difference in the counts of positive and negative tweets per 100,000 residents in the state; for example, in 2013, there was 54 more negative patient experience tweets for every 100,000 Twitter users in the south region. Likewise, there were 28 more negative tweets in the west region compared with the positive tweets. The patient experience tweet rate is the number of patient experience tweets per 100,000 residents in the state [6]; for example, there were 372 patient experience tweets posted in Nevada, 239 in Texas, and 225 in California in 2013 per 100,000 residents.

Overall, the average sentiment polarity shifted to be less negative every year across all the regions in the United States, as shown in Figure 1. The average sentiment polarity rate for the northeast, midwest, south, and west regions shifted from −52, −37, −54, and −27 in 2013 to −36, −17, −33, and −12, respectively, in 2014. The sentiment polarity further shifted toward less negative scores from 2015 to 2016 in all the regions except for the northeast region, which recorded a sentiment polarity rate of −14 in 2015 compared with −17 in 2016.

Similarly, the patient experience tweet rate also decreased across all the states over the 4-year study period. The number of states with at least 200 tweets per 100,000 residents was reduced from 35 states in 2013 to 3 states (Nevada, Oregon, and Alaska) in 2016. The count of patient experience tweets from February 2013 to February 2017 (a total of 49 months) by region is shown in Figure 2. Overall, the south region posted the highest volume of tweets and the northeast posted the lowest volume of tweets during the study period with a visible downward trend across the 4 regions of the United States.

We further examined the negative patient experience tweets with respect to the hour of the day when they were posted. We focused on negative tweets because the average sentiment polarity across all the regions was consistently found to be negative, as shown in Figure 1. Figures 3 and 4 present a set of plots showing the hourly trend and the day-of-week trend respectively for the fraction of the negative patient experience tweets by region for each study year.

The hour-of-day trend revealed that the overall negative tweet fraction exceeded the positive at almost every hour-of-day in all the regions. However, the negative tweet fraction was at its minimum during working hours (8 am-5 pm). The northeast and south regions exhibited very similar tweet patterns during the working hours regardless of the large differences in the tweet counts (Figure 3). The midwest and west regions also show similar patterns to each other. There were similar or higher volumes of positive tweets posted between 10 am and 8 pm in the midwest and west from 2014 to 2016. The fraction of negative tweets was consistently above 0.5 between 10 am and 8 pm in the northeast and south regions.

The day-of-week trend revealed that the overall fraction of negative tweets in all 4 regions was similar over the 4-year study period (Figure 4). The negative tweet fraction was consistently equal to or above 0.5 for all regions in the United States except in the west region in 2015 and 2016. Additionally, Fridays and Saturdays were found to be the least negative days in the week for tweets in the patient experience dataset across all regions and all study years. There was a visible decrease in the negative fraction from Thursday to Friday and a visible increase from Saturday to Sunday in almost all regions every year.

The plots for the hourly and day-of-week tweet counts are shown in Multimedia Appendix 5. We found that the highest number of patient experience tweets was sent from 10 am to 10 pm and on Monday through Thursday across all regions. The south consistently recorded the highest volume of tweets, and the northeast recorded the lowest tweets hourly between 10 am and 10 pm as well as every day of the week. The regional patterns in hourly and day-of-week tweet counts remained similar over the 4-year study period with a visible decrease from 2013 to 2015. Both tweet count trends remained similar across all regions in the years 2015 and 2016.

Population Size Examination

Using the geolocation classifier, we were able to identify whether a tweet was from a metropolitan (≥50,000 persons) area or a nonmetropolitan (<50,000 persons) area. At the national level, we identified 267,894/867,149 tweets in nonmetropolitan areas, accounting for 30.89% of tweets in the geocoded dataset. We identified 599,255/867,149 tweets in metropolitan areas, accounting for 69.11% of tweets in the geocoded dataset. We excluded the tweets from District of Columbia and Puerto Rico for this examination.

Using the sentiment classifier, we observed at the national level that patient experience-related tweets from nonmetropolitan areas had higher negative sentiment when compared with patient experience tweets from metropolitan areas; however, the difference was small (57.3% vs 55.9%). Similarly, we observed patient experience tweets from nonmetropolitan areas to have a slightly lower percentage of positive tweets compared with those tweets from metropolitan areas (42.7% vs 44.1%).

Figure 1. Patient experience tweet sentiment by region over time. K represents thousand, where any number is followed by three zeros (eg, 100K equals 100,000).

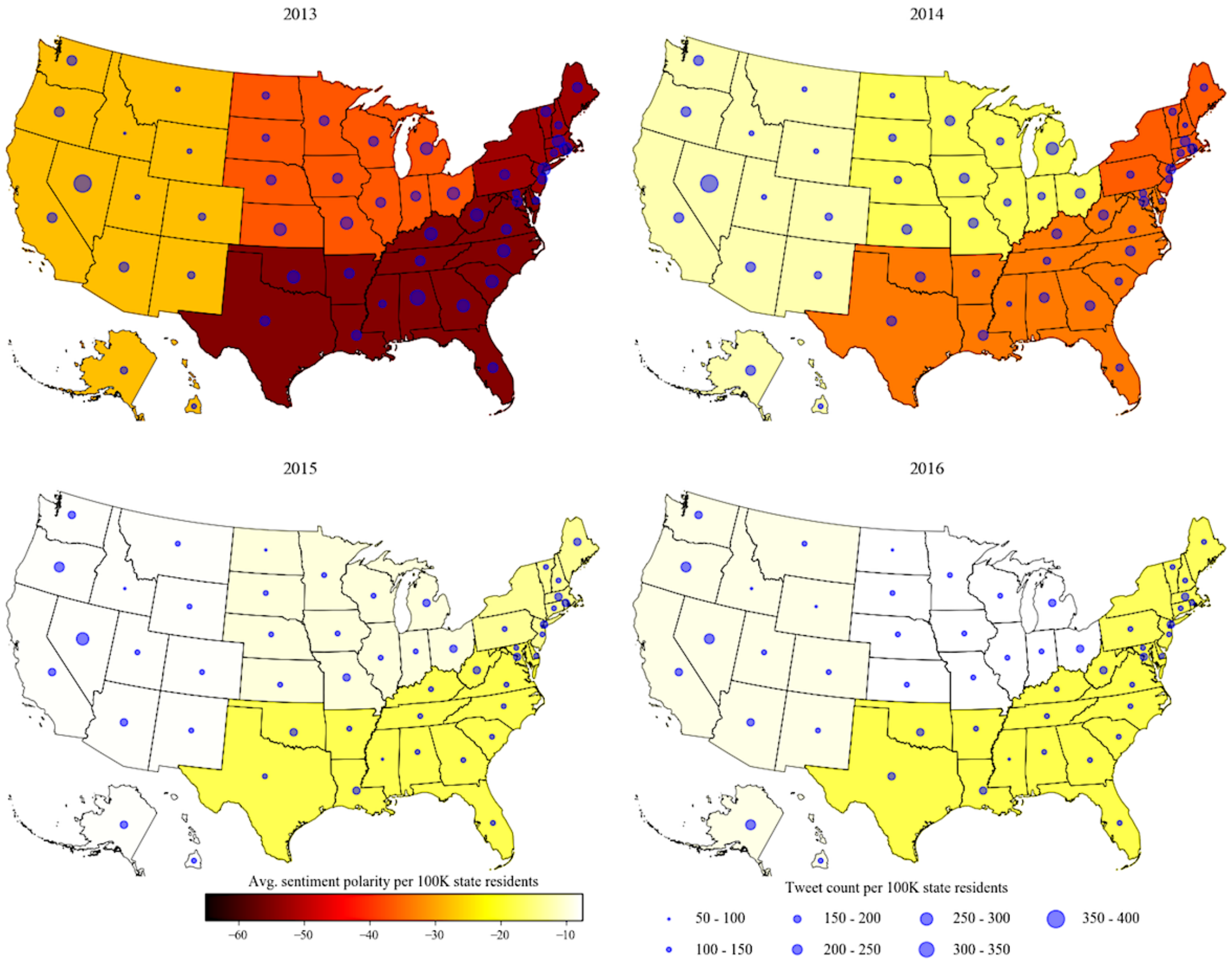


Figure 2. Patient experience tweet volume by region over time. K represents thousand, where any number is followed by three zeros (eg, 100K equals 100,000).

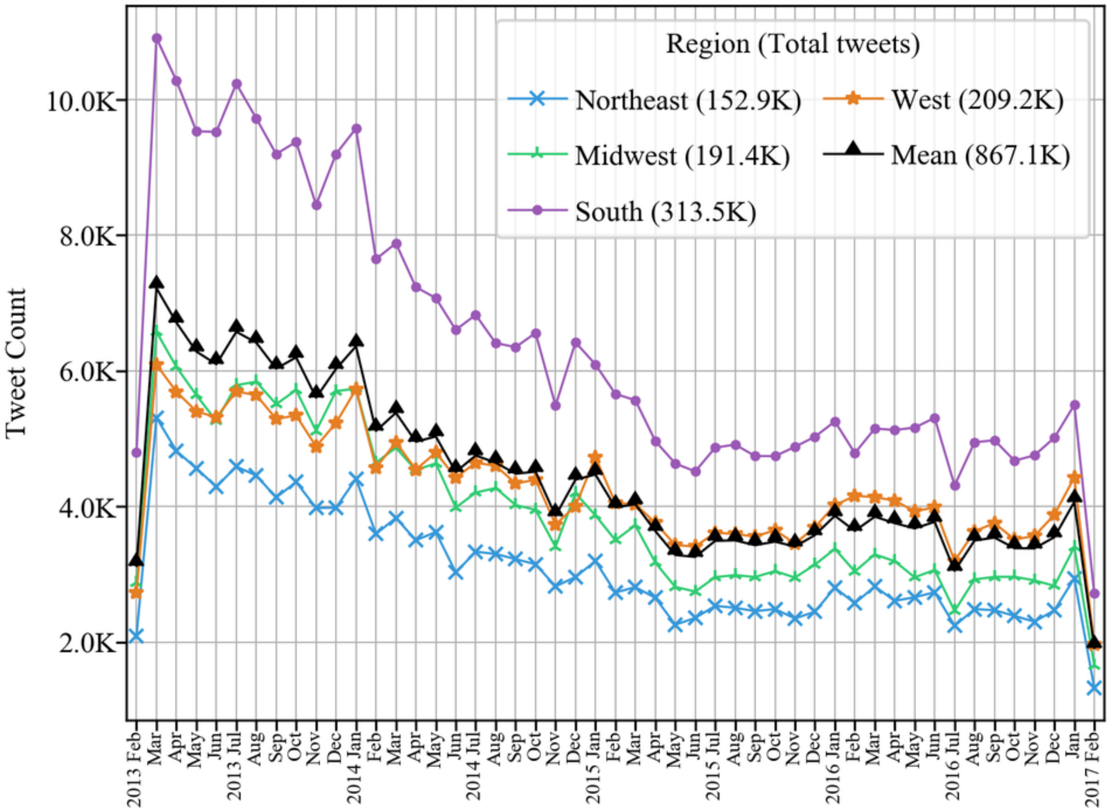


Figure 3. Fraction of negative patient experience tweets by the hour of the day in each region for years 2013-2016.

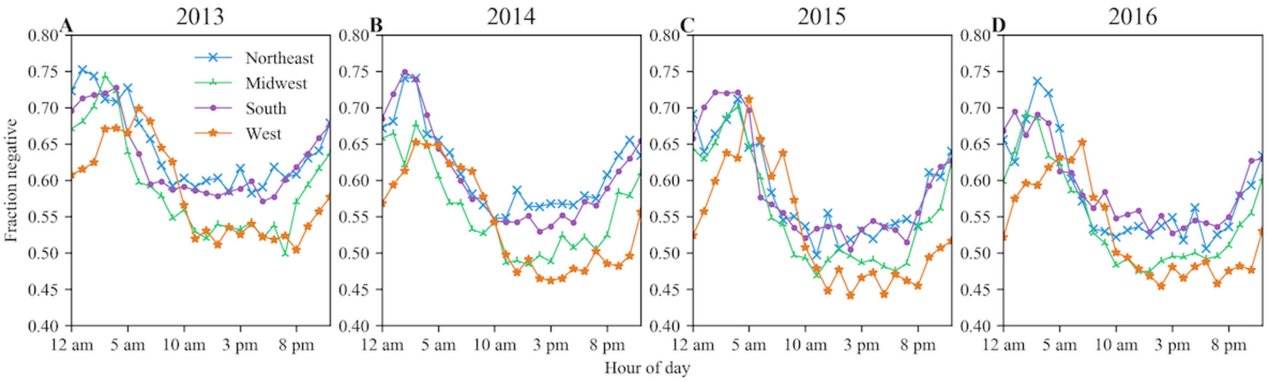
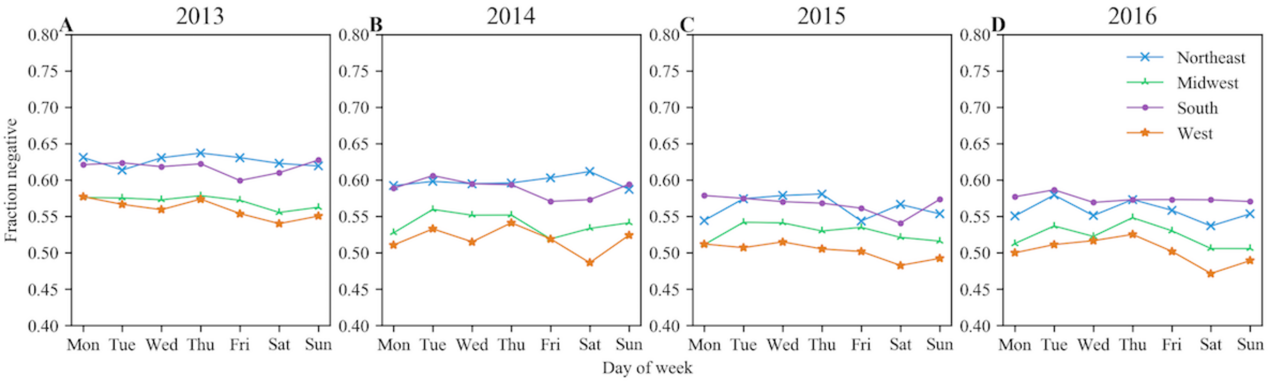


Figure 4. Fraction of negative patient experience tweets by the day of the week in each region for years 2013-2016.



Regionally dividing the metropolitan and nonmetropolitan tweets revealed that the northeast has the largest fraction of metropolitan tweets i.e., 81.12% (124,135/152,944), followed by the west at 73.28% (153,336/209,246), south at 64.65% (202,710/313,543), and midwest at 62.21% (119,074/191,416). The sentiment comparison across all regions and metropolitan or nonmetropolitan areas found that metropolitan patient experience tweets in the western region were most positive (48.3%), and the nonmetropolitan tweets in the south were most negative (60.1%). However, the sentiment percentage difference between the metropolitan and nonmetropolitan tweets within respective regions was also small. The west held the largest difference in sentiment percentage difference with 51.7% negative tweets in metropolitan areas compared with 53.8% in nonmetros. The northeast recorded the smallest sentiment difference (59.2% negative tweets in metropolitan vs 58.4% in nonmetropolitan).

We further divided the metropolitan and nonmetropolitan tweets to study the yearly patterns within and across the regions. In each study year, we found that more negative tweets were posted than positive in all metropolitan and nonmetropolitan areas across all regions. Tweets in the northeast metropolitan area posted the highest percentage of negative tweets (63.0%) across all the regions in 2013. From 2014 to 2016, the southern nonmetropolitan area consistently had the highest percentage of negative tweets with 60.6%, 57.5%, and 58.2% negative tweets for each of these respective study years, 2014, 2015, and 2016, respectively, in the study. However, the western metropolitan and midwestern metropolitan areas recorded the highest and second highest percentage of positive tweets, respectively, each year in the study. The highest positive tweet percentage of the western metropolitan area was 50.5% and the midwestern metropolitan was 49.2%, and both were recorded in 2016. The difference in sentiment percentages within all regions over the 4-year study period was small. The west reported the largest percentage difference in negative sentiments between metropolitan and nonmetropolitan in 2016, 49.5% vs 52.5%, respectively.

Statistical Analysis

In further investigations, we performed statistical tests to identify if there were any significant differences between the sentiment scores of the metropolitan and nonmetropolitan tweets. The shape of score distribution was found to be symmetric bimodal with local maxima on either side of the origin, as seen in Figure 5. Hence, we performed the Mann-Whitney nonparametric test to check if the ranked distribution of the sentiment scores from the metropolitan and the nonmetropolitan areas were approximately equal.

We performed the statistical tests on the sentiment scores at both national and regional level. The sentiment score data were also divided into the following 4 quantiles: Q-1 (0.0, 0.25), Q-2 (0.25, 0.5), Q-3 (0.50, 0.75), and Q-4 (0.75, 1.0) for the analysis. These quantiles represent the relative polarity of the data; for example, the tweets in Q-1 can be viewed as extremely negative compared with the extremely positive tweets in Q-4. Similarly, the tweets in Q-2 and Q-3 can be viewed as mildly negative and mildly positive within a dataset. The descriptive statistics and *P* values of all the statistical tests are shown in Table 2. The table also shows the Cohen *d* effect size for the tests that found significant differences.

The sentiment scores of the metropolitan tweets at the national level were found to be significantly different to the nonmetropolitan tweets (*P*<.001). The sentiment scores of the midwest, south, and west regions' metropolitan tweets were also found to be significantly different from the nonmetropolitan tweets at alpha=0.1%. The *P* value for the northeast region was .003.

After dividing the data into quantiles, the analysis established that the statistical significance could vary at different quantiles and that it was irrespective of the results that we found for the data without dividing it. Nationally, the difference between the metropolitan and the nonmetropolitan tweets was found to be statistically significant for data quantiles Q-1 and Q-3 (*P*<.001).

Figure 5. Sentiment score distribution of all tweets (n=788,904, μ=−0.06, and SD 0.509).

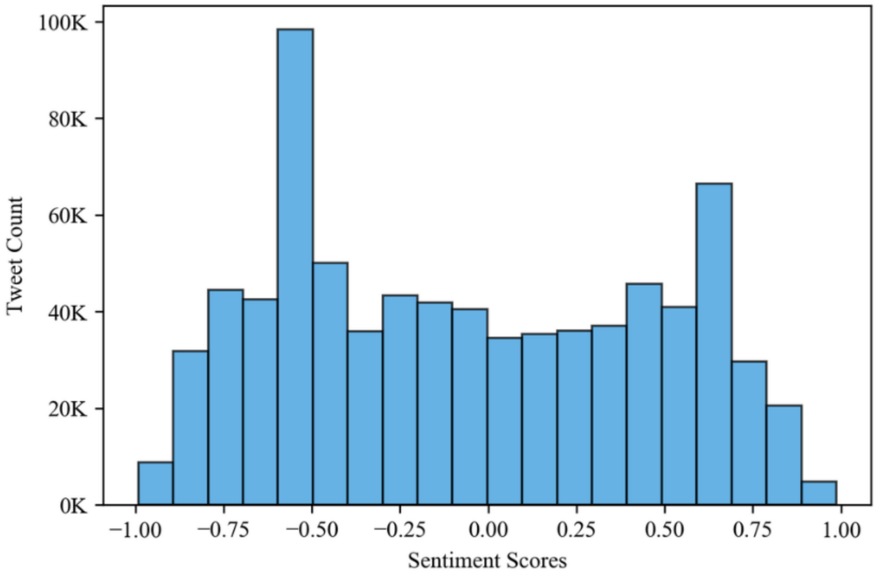


Table 2. National and regional descriptive statistics and nonparametric test results of patient experience tweet sentiments in metropolitan and nonmetropolitan areas.

Tweet region and sentiment score quantiles[a]	Metropolitan tweets		Nonmetropolitan tweets		*P* value	Cohen *d* effect size[b]
	n (%)	Mean (SD)	n (%)	Mean (SD)		
All tweets	544,962 (69.08)	−0.056 (0.511)	243,942 (30.92)	−0.068 (0.505)	<.001	0.023
Q-1	152,796 (68.87)	−0.698 (0.127)	69,067 (31.13)	−0.655 (0.126)	<.001	0.341
Q-2	150,712 (68.36)	−0.351 (0.135)	69,768 (31.64)	−0.349 (0.136)	.002	N/A[c]
Q-3	137,361 (68.94)	0.151 (0.156)	61,876 (31.06)	0.148 (0.156)	<.001	0.019
Q-4	138,339 (70.14)	0.621 (0.128)	58,905 (29.86)	0.620 (0.128)	.04	N/A
Northeast						
All tweets	112,242 (81.08)	−0.086 (0.508)	26,184 (18.92)	0.078 (0.503)	.003	N/A
Q-1	33,560 (81.66)	−0.660 (0.127)	7536 (18.34)	−0.659 (0.127)	.30	N/A
Q-2	32,323 (81.01)	−0.354 (0.134)	7575 (18.99)	−0.345 (0.136)	<.001	0.066
Q-3	27,349 (80.39)	0.150 (0.159)	6671 (19.61)	0.152 (0.160)	.17	N/A
Q-4	26,321 (81.29)	0.619 (0.131)	6059 (18.71)	0.619 (0.128)	.42	N/A
Midwest						
All tweets	108,453 (62.15)	−0.035 (0.512)	66,043 (37.85)	−0.051 (0.508)	<.001	0.032
Q-1	29,045 (61.70)	−0.657 (0.126	18,027 (38.30)	−0.656 (0.126)	.27	N/A
Q-2	29,085 (61.18)	−0.350 (0.136	18,454 (38.82)	−0.347 (0.136)	.12	N/A
Q-3	27,877 (62.44)	0.153 (0.155)	16,770 (37.56)	0.149 (0.156)	.002	N/A
Q-4	28,957 (63.26)	0.623 (0.126)	16,814 (36.74)	0.621 (0.127)	.08	N/A
South						
All tweets	183,829 (64.67)	0.080 (0.506)	100,448 (35.33)	−0.092 (0.50)	<.001	0.024
Q-1	54,537 (64.39)	−0.656 (0.127)	30,163 (35.61)	−0.652 (0.126)	<.001	0.032
Q-2	52,931 (63.93)	−0.353 (0.136)	29,859 (36.07)	−0.353 (0.136)	.27	N/A
Q-3	45,676 (64.39)	0.149 (0.158)	25,263 (35.61)	0.147 (0.158)	.04	N/A
Q-4	43,424 (66.05)	0.617 (0.129)	22,324 (33.95)	0.616 (0.128)	.20	N/A
West						
All tweets	140,438 (73.26)	−0.015 (0.514)	51,267 (26.74)	−0.039 (0.511)	<.001	0.048
Q-1	35,654 (72.77)	−0.660 (0.128)	13,341 (27.23)	−0.659 (0.126)	.44	N/A
Q-2	36,373 (72.38)	−0.345 (0.134)	13,880 (27.62)	−0.346 (0.134)	.02	N/A
Q-3	36,459 (73.46)	0.153 (0.152)	13,172 (26.54)	0.149 (0.154)	.002	N/A
Q-4	39,637 (74.30)	0.626 (0.126)	13,708 (25.70)	0.625 (0.127)	.31	N/A

[a] The 4 sentiment score quantiles are shown using Q-1 (0.0, 0.25), Q-2 (0.25, 0.5), Q-3 (0.50, 0.75), and Q-4 (0.75, 1.0). The results are reported at α=0.1%.

[b] The Cohen *d* effect size was computed for tests that found significant differences.

[c] N/A: not applicable.

This result implies that the extremely negative and mildly positive subset of the metropolitan tweets was significantly different than their counterpart tweets from nonmetropolitan areas at the national level. At the regional level, we found statistically significant differences only for Q-2 in the northeast and Q-1 in the south region. The effect size analysis showed that the metropolitan and nonmetropolitan tweets with extremely negative sentiments (ie, Q-1) had a medium effect size (*d*=0.341) at the national level. The remaining tests showed a low side effect.

Discussion

Principal Findings

Our findings suggest that Twitter is a unique platform for identifying differences in health care and sentiment of discussion across various geographical perspectives over the 4-year study period. The methodologies developed in this study present an informative examination of the sentiments of patient discussions of health care online. By identifying the opinions and attitudes of patients using social media, we can supplement traditional measures of collecting feedback to better understand the care received across the United States. This study has developed or built upon methodologies to examine social data from various geographical perspectives, including national, regional, and population levels across a 4-year study period.

We found that tweets related to patient experience lean toward a higher percent negative sentiment at the national level. Previous research suggests that patient experience scores are directly related to specific factors of care, such as wait time, the quantity of nurses or doctors at the health care facility, or even cost of care [3,4]. Hospital care in the United States has been found to be generally positive [37], and polling measures have found that Americans generally rate their health care experience as good [38,39]. However, this study is not restricted exclusively to hospital data and encompasses a larger scope of care outside of hospitals, which may be attributed to this discrepancy. Web-based examinations of patient experience may differ from what is being reported in interview- and survey-based reports of care. This study also found slight differences in patient experience tweet sentiments that varied owing to region and population. We observed higher percent positive tweets related to patient experiences in the northeast region as well as areas that are defined as metropolitan with a population of ≥50,000 residents. This supports research on the geographic variability of health care cost and outcomes, which can be reflected through Web-based sentiment scores [33,40]. Further research based on these observations can provide insight into the type of care provided in these areas. The sentiment of patient experience tweets in this study over the 4-year study period gradually skews to less negative, which supports previous reports that found that hospital patient experience trends demonstrate positive progress in multi-year evaluations [41].

We observed a downward trend in the tweet volume during the 4-year study period, whereas tweet sentiment was found to increase across all 4 regions of the United States. This trend could be attributed to either a decrease in percent negative tweets posted over time or an increase in percent positive tweets over time for patient experience discussions. Additionally, although Twitter has not publicly commented on this, researchers and developers who work with this platform have observed a decline in Twitter usage in the United States since 2014[42]. This observation may explain why this study also experienced a constant decrease in patient experience tweet count over the 4 study years.

This study provides an in-depth presentation of the time of the day a tweet was posted. We observed the sentiment of tweets to have a lower negative fraction during daytime hours, whereas the sentiment of tweets posted between 8 pm and 10 am tended to have a higher negative fraction. This observation was seen across all 4 regions of the United States. This observation supports previous research that shows that patient care can be compromised during night hours and on weekends particularly because this is a time when facilities may be closed or have reduced staffing [43]; for example, lower survival rates for postcardiac arrest patients have been observed during night and weekend care [44], and measures to improve safety during off-hours care have been recommended [45]. Further research into the significance of these observations is needed.

By examining the differences in tweet sentiment between metropolitan and nonmetropolitan areas, we sought to determine if the discussion of care online differs based on population size. We found that metropolitan areas across the United States have higher percent positive tweets compared with nonmetropolitan areas, which supports research on differences in health care in rural populations compared with the care in urban populations [7]. Metropolitan cities have been found to have better access to care because many have large health care institutions and resources nearby that smaller communities lack [7,8]. Although there are some noted disadvantages to access to care in more populated cities, including longer wait times, travel times, and appointment availability, we would have expected the sentiment of tweets between the metropolitan and nonmetropolitan areas to have a larger difference, which was not observed in this study. Although we do observe statistically significant results in the associations between certain sentiment quantiles and population size based on the metropolitan and nonmetropolitan areas, we do recognize that this is a large-scale dataset and the impact of the results are weak at best. Further research could provide better insight into care expectations and the Web-based conversations between varying population levels in the United States.

Limitations

There were several limitations to our study. First, selection bias could occur from the nature of Twitter usage, a platform which is heavily comprised of adults aged 18 to 29 [46]. Representation of tweets may not be evenly distributed across all age groups. Second, we collected our data based on selected keywords related to patient experience, which may not have captured all tweets on the subject matter. Owing to the broad nature of the intended dataset, there is a chance some discussions of patient care were missed. There are limitations to the selected classifier for identifying patient experience tweets as well. We found that the selected classifier achieved an accuracy of 83% with the precision and recall of 70% and 69%, respectively, for the patient experience tweet class, which suggests there is still a chance that tweets discuss health care, but perhaps a tweet that is not an exposure to health care could be captured in this dataset. As previously noted, we observed a decrease in the tweet volume over the 4-study year period, which could indicate that people are posting less on Twitter over time. To minimize the bias of the tweet count, we normalized the patient experience data using yearly state population estimates or yearly tweet counts in our analysis. We present the count data as supplemental information in the analysis. The effect of decreasing tweet counts may introduce bias in the observed data, and this needs to be explored further.

Additionally, there are limitations to state identification because the human validation of our geocoding engine has 91% accuracy. Therefore, there is a 9% chance of error in the inferred states. The errors in the geolocation are primarily owing to the way users provide their location information in their profile;

for example, if a Twitter user provides only a city name with no state or country information included in the location field in her profile, the inferred state might be incorrect. Furthermore, querying any location in our engine produces a list of possible options for the state and country. However, we can only select 1 out of all potential options. We are currently choosing the first one on the list.

Finally, our location engine, which infers the state and if a tweet is from a metropolitan or nonmetropolitan area, is based on boundary polygon data on state and urban areas provided by the US Census Bureau. Although we used the highest available data resolution, the inferred location might be incorrect if the tweet geo-coordinates fall very close to the polygon boundaries. There is also a limitation of using a denominator of "per 100,000 residents per state" for our understanding of this dataset. We acknowledge that this denominator represents neither patients on Twitter per state nor Twitter users per state. We used this denominator as referenced by state population US Census Bureau data.

There are ethical considerations that must be considered when using data from social media sites such as Twitter. Understandably, users have concerns about privacy and confidentiality of information posted online. Interestingly, Web-based data used for social benefit or public health interest are often perceived to be more acceptable in social media research among users [47,48]. Even though we acknowledge the concerns of users, all information used in this study was acquired for academic research and was restricted to publicly available posts that users have selected not to post privately. Additionally, we attempted to address concerns about privacy and confidentiality by analyzing and disseminating aggregated numerical data only.

Future Direction

The findings of this study have implications for future research examining patient feedback online and the usefulness of the knowledge it can provide. Twitter can be prospectively or historically monitored by geographical location to determine how patients feel about the care they receive. This novel approach presents patients with the opportunity to freely discuss their feedback on all aspects of care provided without being limited to the restrictions from more traditional structured questionnaires. Although a user-based approach was outside the scope of this study, future research using the methodologies presented could consider analyzing user-specific data to further examine geographical and temporal differences in patient experience discussions. Additionally, Twitter surveillance of Web-based discussions may provide health care providers, health institutions, and policy makers with both positive and negative trends in the care received in their jurisdiction. This can inform stakeholders of where health care can be improved, particularly during a time when the influence of patient engagement can direct where limited resources should be allocated. Furthermore, these data have the power to provide future research into differences of patient feedback between population demographics, topics of discussion, or even questions to understand if patients are receiving the right care at the right time. Deeper knowledge on the discussions of care online can provide valuable and insightful information, which has the power to influence how health care is provided across the United States.

Conclusions

This study presents methodologies for a deeper understanding of Web-based discussion related to patient experience across space and time. Twitter, as a social media platform, provides a unique and unsolicited perspective from users. This characterization of data provides a unique opportunity to examine geographic and temporal differences in the sentiments of patient opinions and feedback. The findings provided in this study can lead to further research and understanding of the culture of health in the United States as provided by real-time social data.

Acknowledgments

KCS and GT contributed to study inception, data collection, data analysis and interpretation, and drafting and critical revision of the manuscript. YH contributed to data analysis and interpretation and critical revision of the manuscript. JSB and JBH contributed to study inception, data collection, data analysis and interpretation, and critical revision of the manuscript. This study was funded by the Robert Wood Johnson Foundation Grant. The funder played no role in the study design; collection, analysis, or interpretation of data; writing of the manuscript; or decision to submit the manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Keyword classes, keywords, and example rules that were used to extract the patient experience Twitter data.

[PDF File (Adobe PDF File), 42KB - jmir_v20i10e10043_app1.pdf]

Multimedia Appendix 2

Amazon mechanical turk curation guide.

[PDF File (Adobe PDF File), 33KB - jmir_v20i10e10043_app2.pdf]

Multimedia Appendix 3

Tweet classification.

[PDF File (Adobe PDF File), 36KB - jmir_v20i10e10043_app3.pdf]

Multimedia Appendix 4

Tweet extraction, curation, and geolocation flowchart.

[PDF File (Adobe PDF File), 176KB - jmir_v20i10e10043_app4.pdf]

Multimedia Appendix 5

Patient experience tweet counts by the hour-of-day and day-of-week for each US region from years 2013 to 2016.

[PDF File (Adobe PDF File), 160KB - jmir_v20i10e10043_app5.pdf]

References

1. Charmel PA, Frampton SB. Building the business case for patient-centered care. Healthc Financ Manage 2008 Mar;62(3):80-85. [Medline: 19097611]
2. Centers for Medicare & Medicaid Services. 2017. HCAHPS: Patients' Perspectives of Care Survey URL: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/HospitalQualityInits/HospitalHCAHPS.html [accessed 2018-09-24] [WebCite Cache ID 72fql9Fpb]
3. Anhang PR, Elliott MN, Zaslavsky AM, Hays RD, Lehrman WG, Rybowski L, et al. Examining the role of patient experience surveys in measuring health care quality. Med Care Res Rev 2014 Oct;71(5):522-554 [FREE Full text] [doi: 10.1177/1077558714541480] [Medline: 25027409]
4. Doyle C, Lennox L, Bell D. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. BMJ Open 2013 Jan 03;3(1):piie001570 [FREE Full text] [doi: 10.1136/bmjopen-2012-001570] [Medline: 23293244]
5. Matthews KA, Croft JB, Liu Y, Lu H, Kanny D, Wheaton AG, et al. Health-Related Behaviors by Urban-Rural County Classification - United States, 2013. MMWR Surveill Summ 2017 Dec 03;66(5):1-8 [FREE Full text] [doi: 10.15585/mmwr.ss6605a1] [Medline: 28151923]
6. Hall SA, Kaufman JS, Ricketts TC. Defining urban and rural areas in U.S. epidemiologic studies. J Urban Health 2006 Mar;83(2):162-175 [FREE Full text] [doi: 10.1007/s11524-005-9016-3] [Medline: 16736366]
7. Spasojevic N, Vasilj I, Hrabac B, Celik D. Rural - Urban Differences In Health Care Quality Assessment. Mater Sociomed 2015 Dec;27(6):409-411 [FREE Full text] [doi: 10.5455/msm.2015.27.409-411] [Medline: 26937222]
8. Douthit N, Kiv S, Dwolatzky T, Biswas S. Exposing some important barriers to health care access in the rural USA. Public Health 2015 Jun;129(6):611-620. [doi: 10.1016/j.puhe.2015.04.001] [Medline: 26025176]
9. Verhoef LM, Van de Belt TH, Engelen LJLPG, 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]
10. Mazor KM, Clauser BE, Field T, Yood RA, Gurwitz JH. A demonstration of the impact of response bias on the results of patient satisfaction surveys. Health Serv Res 2002 Oct;37(5):1403-1417 [FREE Full text] [Medline: 12479503]
11. Strauss M. Pew Research Center. Most patients in US have high praise for their health care providers URL: http://www.pewresearch.org/fact-tank/2017/08/02/most-patients-in-u-s-have-high-praise-for-their-health-care-providers [accessed 2018-02-26] [WebCite Cache ID 6xWQyVvXh]
12. Rainie L, Funk C. Pew Research Center: Internet, Science & Tech. 2015 Jan 29. Chapter 2: Perspectives on the Place of Science in Society URL: http://www.pewinternet.org/2015/01/29/chapter-2-perspectives-on-the-place-of-science-in-society/ [accessed 2018-02-26] [WebCite Cache ID 6xWRO9qY7]
13. The Statistics Portal. 2018. Twitter: Number of monthly active U.S. users 2010-2018 URL: https://www.statista.com/statistics/274564/monthly-active-twitter-users-in-the-united-states/ [accessed 2018-07-03] [WebCite Cache ID 70n0LfRep]
14. Hawkins JB, Brownstein JS, Tuli G, Runels T, Broecker K, Nsoesie EO, et al. Measuring patient-perceived quality of care in US hospitals using Twitter. BMJ Qual Saf 2016 Dec;25(6):404-413 [FREE Full text] [doi: 10.1136/bmjqs-2015-004309] [Medline: 26464518]
15. Tighe PJ, Goldsmith RC, Gravenstein M, Bernard HR, Fillingim RB. The painful tweet: text, sentiment, and community structure analyses of tweets pertaining to pain. J Med Internet Res 2015;17(4):e84 [FREE Full text] [doi: 10.2196/jmir.3769] [Medline: 25843553]

16. Downing NS, Cloninger A, Venkatesh AK, Hsieh A, Drye EE, Coifman RR, et al. Describing the performance of U.S. hospitals by applying big data analytics. PLoS One 2017 Jun;12(6):e0179603 [FREE Full text] [doi: 10.1371/journal.pone.0179603] [Medline: 28662045]
17. Gohil S, Vuik S, Darzi A. Sentiment Analysis of Health Care Tweets: Review of the Methods Used. JMIR Public Health Surveill 2018 Apr 23;4(2):e43 [FREE Full text] [doi: 10.2196/publichealth.5789] [Medline: 29685871]
18. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Use of sentiment analysis for capturing patient experience from free-text comments posted online. J Med Internet Res 2013;15(11):e239 [FREE Full text] [doi: 10.2196/jmir.2721] [Medline: 24184993]
19. 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;16(3):e79 [FREE Full text] [doi: 10.2196/jmir.2679] [Medline: 24642014]
20. Privacy Policy. 2018 May 25. Twitter URL: https://twitter.com/en/privacy [accessed 2018-07-17] [WebCite Cache ID 70zE4CiGv]
21. Get Better Results with Amazon Mechanical Turk Masters. 2011 Jun 23. Amazon Web Services URL: https://aws.amazon.com/blogs/aws/amazon-mechanical-turk-master-workers/ [accessed 2018-07-03] [WebCite Cache ID 70iUxUxk3]
22. Colditz JB, Chu K, Emery SL, Larkin CR, James AE, Welling J, et al. Toward Real-Time Infoveillance of Twitter Health Messages. Am J Public Health 2018 Aug;108(8):1009-1014. [doi: 10.2105/AJPH.2018.304497] [Medline: 29927648]
23. Graham M, Hale SA, Gaffney D. Where in the World Are You? Geolocation and Language Identification in Twitter. The Professional Geographer 2014 May 19;66(4):568-578. [doi: 10.1080/00330124.2014.907699]
24. Google. Developer Guide. Geocoding API URL: https://developers.google.com/maps/documentation/geocoding/intro [accessed 2018-02-23] [WebCite Cache ID 6xWRyUI7F]
25. US Census Bureau. Cartographic Boundary Shapefiles - States URL: https://www.census.gov/geo/maps-data/data/cbf/cbf_state.html [accessed 2018-09-24] [WebCite Cache ID 72frcdTkc]
26. Google. Geocoding Addresses Best Practices. Geocoding API URL: https://developers.google.com/maps/documentation/geocoding/best-practices [accessed 2018-07-10] [WebCite Cache ID 710cT2tZG]
27. Gianfranco G, Rinnone F. Online Geocoding Services: A Benchmarking Analysis to Some European Cities. In: Baltic Geodetic Congress.: IEEE; 2017 Presented at: BGC Geomatics; June 22-25 2017; Gdansk, Poland p. 273-281. [doi: 10.1109/BGC.Geomatics.2017.12]
28. Ahlers D, Boll S. University of Oldenburg, Germany. 2009. On the Accuracy of Online Geocoders URL: http://www.informatik.uni-oldenburg.de/~there/geoinformatik09.pdf [accessed 2018-07-18] [WebCite Cache ID 710dH2Pzq]
29. Clemens K. Geocoding with openstreetmap data. In: GEOProcessing 201.: IARIA; 2015 Presented at: GEOProcessing 2015: The Seventh International Conference on Advanced Geographic Information Systems, Applications, and Services; February 22-27, 2015; Lisbon, Portugal p. 1-2.
30. Dredze M, Paul M, Bergsma S, Tran H. Carmen: A Twitter geolocation system with applications to public health. In: Workshop on Expanding the Boundaries of Health Informatics Using Artificial Intelligence. 2013 Presented at: AAAI Conference on Artificial Intelligence; July 14-18, 2013; Bellevue, Washington, USA p. 20-24.
31. US Census Bureau. 2018. Geography: Regions URL: https://www.census.gov/geo/reference/webatlas/regions.html [accessed 2018-08-02] [WebCite Cache ID 72fspWZTS]
32. US Census Bureau. 2018. Statistical Groupings of States and Counties URL: https://www2.census.gov/geo/pdfs/reference/GARM/Ch6GARM.pdf [accessed 2018-09-24] [WebCite Cache ID 72fsvp66J]
33. Rosenberg BL, Kellar JA, Labno A, Matheson DHM, Ringel M, VonAchen P, et al. Quantifying Geographic Variation in Health Care Outcomes in the United States before and after Risk-Adjustment. PLoS One 2016;11(12):e0166762 [FREE Full text] [doi: 10.1371/journal.pone.0166762] [Medline: 27973617]
34. Gilbert E, Hutto CJ. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. 2014 Presented at: Eighth International AAAI Conference on Weblogs and Social Media; June 2-4, 2014; Ann Arbor, Michigan p. 216-225 URL: http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/download/8109/8122
35. Kralj NP, Smailović J, Sluban B, Mozetič I. Sentiment of Emojis. PLoS One 2015 Dec;10(12):e0144296 [FREE Full text] [doi: 10.1371/journal.pone.0144296] [Medline: 26641093]
36. US Census Bureau. 2012 Sep 1. Geography: Cartographic Boundary Shapefiles - Urban Areas URL: https://www.census.gov/geo/maps-data/data/cbf/cbf_ua.html [accessed 2018-09-24] [WebCite Cache ID 72ftHqgs1]
37. HCAHPS. 2018. Summary Analyses URL: http://www.hcahpsonline.org/en/summary-analyses/ [accessed 2018-07-03] [WebCite Cache ID 70waIyZAP]
38. NPR. 2016. Patient' Perspectives on Healthcare in the United States URL: https://www.npr.org/assets/img/2016/02/26/PatientPerspectives.pdf [accessed 2018-07-03] [WebCite Cache ID 70waYK36j]
39. Gallup News. 2017 Nov 30. U.S. Healthcare Quality Ratings Among Lowest Since '12 URL: https://tinyurl.com/y7ly8tm3 [accessed 2018-07-03] [WebCite Cache ID 70waijm93]
40. Newhouse J, Garber A, Graham R, Mccoy M, Mancher M, Kibria A, editors. Variation in Health Care Spending: Target Decision Making, Not Geography. In: Institute of Medicine. Washington, DC: The National Academies Press; Oct 1, 2013.

41. Wolf J. The Beryl Institute. 2013. The State of Patient Experience in American Hospitals 2013: Positive trends and opportunities for the future URL: https://c.ymcdn.com/sites/www.theberylinstitute.org/resource/resmgr/benchmarkingpaper_2013.pdf [accessed 2018-07-03] [WebCite Cache ID 70wbCKxUt]
42. Edwards J. Business Insider. 2016 Feb 02. Leaked Twitter API data shows the number of tweets is in serious decline URL: http://www.businessinsider.com/tweets-on-twitter-is-in-serious-decline-2016-2 [accessed 2018-03-09] [WebCite Cache ID 710McFDTZ]
43. Wong HJ, Morra D. Excellent hospital care for all: open and operating 24/7. J Gen Intern Med 2011 Sep;26(9):1050-1052 [FREE Full text] [doi: 10.1007/s11606-011-1715-8] [Medline: 21499824]
44. Ofoma UR, Basnet S, Berger A, Kirchner HL, Girotra S, American Heart Association Get With the Guidelines – Resuscitation Investigators. Trends in Survival After In-Hospital Cardiac & Arrest During Nights and Weekends. J Am Coll Cardiol 2018 Jan 30;71(4):402-411 [FREE Full text] [doi: 10.1016/j.jacc.2017.11.043] [Medline: 29389356]
45. Shulkin DJ. Assessing hospital safety on nights and weekends: the SWAN tool. J Patient Saf 2009 Jun;5(2):75-78. [doi: 10.1097/PTS.0b013e3181a5db10] [Medline: 19920445]
46. Pew Research Center: Internet, Science & Tech. 2017 Jan 12. Social Media Fact Sheet URL: http://www.pewinternet.org/fact-sheet/social-media/ [accessed 2018-01-26] [WebCite Cache ID 6xWRaajnX]
47. 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]
48. Conway M. Ethical issues in using Twitter for public health surveillance and research: developing a taxonomy of ethical concepts from the research literature. J Med Internet Res 2014;16(12):e290 [FREE Full text] [doi: 10.2196/jmir.3617] [Medline: 25533619]

Abbreviations

GPS: Global Positioning System
MTurk: Amazon Mechanical Turk
NLP: natural language processing
VADER: Valence Aware Dictionary for Sentiment Reasoning

Edited by G Eysenbach; submitted 09.03.18; peer-reviewed by D Mciver, A Cyr, L McCann, J Colditz, Z Tingting; comments to author 09.04.18; revised version received 18.07.18; accepted 07.08.18; published 12.10.18.

Please cite as:
Sewalk KC, Tuli G, Hswen Y, Brownstein JS, Hawkins JB
Using Twitter to Examine Web-Based Patient Experience Sentiments in the United States: Longitudinal Study
J Med Internet Res 2018;20(10):e10043
URL: http://www.jmir.org/2018/10/e10043/
doi:10.2196/10043
PMID:30314959

Original Paper

Identification of Primary Medication Concerns Regarding Thyroid Hormone Replacement Therapy From Online Patient Medication Reviews: Text Mining of Social Network Data

So Hyun Park[1,2*], PharmD; Song Hee Hong[1*], PhD

[1]Research Institute of Pharmaceutical Science, College of Pharmacy, Seoul National University, Seoul, Republic Of Korea

[2]Korea Institute of Drug Safety and Risk Management, Drug Safety Information Office of Adverse Drug Reaction Relief, Anyang, Republic Of Korea

[*]all authors contributed equally

Corresponding Author:
Song Hee Hong, PhD
Research Institute of Pharmaceutical Science
College of Pharmacy
Seoul National University
Gwanak-gu Gwanak-ro 1
Seoul,
Republic Of Korea
Phone: 82 02 880 1547
Email: songhhong@snu.ac.kr

Abstract

Background: Patients with hypothyroidism report poor health-related quality of life despite having undergone thyroid hormone replacement therapy (THRT). Understanding patient concerns regarding levothyroxine can help improve the treatment outcomes of THRT.

Objective: This study aimed to (1) identify the distinctive themes in patient concerns regarding THRT, (2) determine whether patients have unique primary medication concerns specific to their demographics, and (3) determine the predictability of primary medication concerns on patient treatment satisfaction.

Methods: We collected patient reviews from WebMD in the United States (1037 reviews about generic levothyroxine and 1075 reviews about the brand version) posted between September 1, 2007, and January 30, 2017. We used natural language processing to identify the themes of medication concerns. Multiple regression analyses were conducted in order to examine the predictability of the primary medication concerns on patient treatment satisfaction.

Results: Natural language processing of the patient reviews of levothyroxine posted on a social networking site produced 6 distinctive themes of patient medication concerns related to levothyroxine treatment: how to take the drug, treatment initiation, dose adjustment, symptoms of pain, generic substitutability, and appearance. Patients had different primary medication concerns unique to their gender, age, and treatment duration. Furthermore, treatment satisfaction on levothyroxine depended on what primary medication concerns the patient had.

Conclusions: Natural language processing of text content available on social media could identify different themes of patient medication concerns that can be validated in future studies to inform the design of tailored medication counseling for improved patient treatment satisfaction.

(J Med Internet Res 2018;20(10):e11085) doi:10.2196/11085

KEYWORDS

medication counseling; social network data; primary medication concerns; satisfaction with levothyroxine treatment

Introduction

Patients with hypothyroidism are treated with the most frequently prescribed drug in the United States, levothyroxine [1], but report poor quality of life despite the treatment [2-4]. Poor quality of life while receiving the treatment possibly reflects the different views that clinicians and patients hold toward the treatment success of thyroid hormone replacement therapy (THRT). Clinicians judge the treatment success based on a restored reference range of thyroid hormone in the blood. On the other hand, patients feel that the treatment is successful when they are free of symptoms such as fatigue, depression, and muscle cramps that are specific to hypothyroidism. Historically, the patient view has not been given adequate attention. Now, scientific evidence has grown for the health care community to embrace the patient-centric paradigm. Specifically for THRT, endocrinologists have come to accept the patient view following a series of studies that reported that some patients do not in fact get better on THRT, despite having a reference range of the thyroid hormone in the bloodstream. For this reason, European and American thyroid professional associations now acknowledge that patients who do not thrive on levothyroxine (T4) alone may try a combination therapy of T4 and T3 (liothyronine) [5,6]. Armour, a desiccated natural thyroid hormone, offers such a combination. A synthetic version also exists with a 4:1 mix of T4 and T3 and a more consistent composition than Armour [6,7]. However, routine use of these combinations is not recommended [8].

Studies pertaining to the patient view have mostly relied on questionnaire-based surveys that employ structured measurements of patient-reported outcomes [9-11]. Perhaps a better way to understand the patient view would be to explore an alternative source of data. Patients have begun to report their health care experiences on social networks [11]. As these reports come directly from patients with no filtering in between, they have the potential to genuinely represent patient experiences. Furthermore, they continuously accumulate on a large scale every day. A Pew Research study in 2012 reported that 61% of adults have read someone else’s health experience on a social network [12]. Many health websites such as WebMD and Ask-a-Patient provide open forums for patients to post ratings and reviews regarding their medication experience [13]. These online communities have become a valuable source of information for the study of patient medication experience [14].

Recent advancement in natural language processing (NLP) has enabled researchers to analyze unstructured text data posted by patients on social networks. NLP of text content on health-related social media has helped identify myriad health topics across a wide range of demographics [15]. Specifically in the area of THRT, NLP has been used to compare the patient ratings of levothyroxine and Armour along with their respective performance potentials [13]; the authors of the comparative study had recognized that patients were dissatisfied with levothyroxine and thus aimed to document the potential superiority of Armour [13]. However, before steering patients away from levothyroxine, a study needs to be conducted fully documenting the concerns of patients being treated with levothyroxine. Our study aimed to identify the distinctive themes of patient concerns regarding levothyroxine treatment based on an alternative data source in the form of a social network to inform the importance of patient-tailored medication counseling for THRT.

The specific aims of this study were to (1) identify the distinctive themes of patient concerns regarding THRT; (2) determine whether patients have unique primary medication concerns specific to their age, gender, and treatment duration; and (3) determine the predictability of primary medication concern on patient satisfaction with THRT, controlling for age, gender, treatment duration, and rating year. The results of this study can inform the design of a patient-centric medication therapy management that accounts for concerns unique to individual patients.

Methods

Study Design and Settings

Data for this study came from patient reviews and ratings of levothyroxine posted on a social network site (WebMD) in the United States between September 1, 2007, and January 30, 2017. WebMD runs a social network service that allows patients to browse patient reviews of prescription drugs based on their medication experience and to post their own reviews. Patients wishing to post their own reviews choose ratings on a scale of 1 to 5 stars for effectiveness, ease of use, and drug satisfaction and choose a reason for taking the drug. Patients can also answer questions relating to age, sex, and treatment duration. In addition, they can provide an open-ended comment. Finally, patients are required to enter the letters of a Captcha-like picture visible on the screen to prevent illegitimate use by automated programs and indicate that they agree to abide by the WebMD terms and conditions and privacy policy.

Text Mining

Text mining required two preprocesses for building text corpora from the patient reviews. The first preprocess was cleaning, through which stop-words such as articles, numbers, punctuations, and demonstrative pronouns were removed from the text. Additionally frequently appearing but irrelevant words such as “take,” “im,” “cant,” “dont,” “thyroid,” “taken,” “t,” “now,” “takeing,” “takes,” “taking,” “age,” “ive,” “also,” “almost,” “els,” “else,” “far,” and “since” were also removed. The second preprocess was stemming, which reduced the words to their “stems.” Finally, a document-term matrix consisting of words along with their frequencies appearing in all the reviews was constructed. The frequency of each word was calculated from the term frequency-inverse document frequency [16,17]. For conducting the data preprocesses and analyses, open source R version 3.3.2 (The R Foundation) was used.

NLP was used to automate the task of information retrieval, analysis, and prediction inherent in languages used in patient reviews. Specifically, basic NLP methods such as tokenization, stop words, and stemming were used to process the content of patient reviews using the R software.

Latent Dirichlet allocation (LDA) was used to automatically discover hidden topics from a set of patient reviews, each of which contained a bag of words. The algorithm treated each review as a mixture of several topics and each topic as a distribution of words. By understanding topic and word distributions among the patient reviews, hidden information in the text could be found automatically [18]. For the LDA topic modeling, the numbers to be set of initial samples to be discarded, sampling iterations, and topics were 4000, 2000, and 6, respectively. The LDA package of the open source R was used as the analysis tool.

Statistical Analysis

The frequency of each theme of patient medication concern was computed for each review and then summed for all reviews. The chi-square test was performed to test whether the distribution of themes of patient medication concerns varied with age, sex, and treatment duration. The percent distribution of themes of medication concerns was also computed for each subgroup of age, gender, and treatment duration along with their standard deviations. Statistical significance was tested at an alpha of .05. To determine the predictability of primary medication concern on treatment satisfaction, multiple linear regression analysis was conducted, controlling for age, sex, treatment duration, and the year the rating was posted.

Results

Description of Patient Medication Reviews

The total number of patient medication reviews and ratings posted on WebMD collected for this study was 2112, which included 1037 reviews of levothyroxine and 1075 reviews of its brand version, Synthroid. After eliminating those that had no comments, the number decreased to 1768 (Table 1). While the reviews were mostly posted by patients themselves (1694/1768, 95.81%), in rare cases (27/1768, 1.53%) they were posted by caregivers. Interestingly, almost half of the reviews (772/1768, 43.67%) were posted by those who had taken the drug for less than a year. Regarding the reason for taking the drug, underactive thyroid (1300/1768, 73.53%) was the most common, followed by thyroid cancer (115/1768, 6.50%). Females (1485/1768, 83.99%) exceeded males (214/1768, 12.1%) with respect to posting of reviews. As for the age of the reviewers, the 45- to 64-year-old group was the most dominant (863/1768, 48.81%). The reviews were posted between 2007 and 2017 with the largest number (351/1768, 19.85%) posted in 2009, followed by 2010 (321/1768, 18.16%) and 2011 (234/1768, 13.24%). Since 2011, posts had a steadily declining trend.

Identification of Distinctive Themes of Patient Medication Concerns

The LDA topic modeling of the patient reviews produced 6 distinctive topics, each of which captured a hidden theme of patient medication concern (Table 2). Each hidden theme identified was given a name representing a certain aspect of medication treatment based on the top 20 most frequently appearing words [19]. A gallery of word clouds lists a complete set of words contained in each theme (Multimedia Appendix 1). The theme of patient medication concern about how to take the drug was named based on the keywords “work,” “time,” “hour,” “dose,” “water,” and “morning.” This concern had the least frequency share (300/2194, 13.67%); that is, it appeared least frequently out of the 6 themes of medication concerns. Likewise, the remaining 5 themes of patient medication concerns (treatment initiation, dose adjustment, symptoms of pain, generic substitutability, and appearance) were identified from scouting the list of top 20 keywords. The appearance theme of patient medication concern was named after scouting the keywords of “weight gain,” “hair loss,” and “dried skin.” The theme of appearance arose most frequently (436/2194, 19.87%), followed by dose adjustment (396/2194, 18.05%) and symptoms of pain (379/2194, 17.27%). The dose adjustment theme had keywords like “mcg,” “dosage,” “increase,” “change,” and “removal” in the list of top 20 words. The symptoms of pain theme had words like “severe,” “bad,” “pain,” “symptom,” “headache,” “heart,” “bodies,” “muscle,” “fatigue,” and “leg” in the list. Treatment initiation had keywords like “feel,” “just,” “better,” “start,” “sleep,” “first,” and “time” in the list. Generic substitutability had keywords like “synthroid,” “generic,” brand,” “blood,” and “test.”

Uniqueness of Primary Medication Concerns Among Different Patients

It was of interest to determine whether patients had unique medication concerns regarding levothyroxine treatment specific to their gender, age, and treatment duration. First, for gender, female patients mentioned “appearance” (mean 21.18% [SD 0.41]) as the most frequent concern (or primary medication concern) and how to take the drug (mean 12.69% [SD 0.33]) as the least frequent concern (Figure 1). However, male patients were least concerned about appearance (mean 13.06% [SD 0.34]) and most concerned about symptoms of pain (mean 19.93% [SD 0.40]).

Patient medication concerns regarding dose adjustment, symptoms of pain, and appearance were quite different according to age. The youngest (13 to 44 years) were the most concerned about dose adjustment (mean 21.33% [SD 0.41]) and appearance (mean 21.03% [SD 0.41]) while the oldest (65 years and older) were most concerned about symptoms of pain (mean 22.39% [SD 0.42]). Patient medication concerns were also quite different with treatment duration. Patients on levothyroxine less than a year were more concerned about appearance (mean 21.64% [SD 0.41]) and drug therapy initiation (mean 19.33% [SD 0.39]) while patients on levothyroxine 5 years or longer were more concerned about how to take the drug (mean 18.96% [SD 0.39]), dose adjustment (mean 20.70% [SD 0.41]), and generic substitutability (mean 19.75% [SD 0.40]).

Table 1. Characteristics of patient medication review writers.

Demographics	Total (n=1768), n (%)
Gender	
Male	214 (12.10)
Female	1485 (83.99)
N/A	69 (3.90)
Age, years	
13-44	562 (31.79)
45-64	863 (48.81)
65 and over	306 (17.31)
N/A	37 (2.09)
Treatment duration, years	
Less than 1	772 (43.67)
1-5	423 (23.93)
More than 5	505 (28.56)
N/A	68 (3.85)
Reasons of taking levothyroxine	
Progressive disease of the thyroid gland	56 (3.17)
Additional treatment for thyroid cancer	1 (0.06)
Decreased thyroid function existing from birth	20 (1.13)
Enlarged thyroid gland	42 (2.38)
Myxedema coma	1 (0.06)
Serious decrease in thyroid function	81 (4.58)
Thyroid cancer	115 (6.50)
Underactive thyroid	1300 (73.53)
Other	152 (8.60)
Reviewer type	
Caregiver	27 (1.53)
Patient	1694 (95.81)
N/A	42 (2.66)
Year	
2007	73 (4.13)
2008	210 (11.88)
2009	351 (19.85)
2010	321 (18.16)
2011	234 (13.24)
2012	178 (10.07)
2013	197 (11.14)
2014	107 (6.05)
2015	39 (2.21)
2016	54 (3.05)
2017	4 (0.23)

Table 2. Identification of 6 distinctive themes of patient medication concerns about levothyroxine from patient reviews.

Theme name	Top 20 words	Frequency (n=2194[a]), n (%)
How to take the drug	work, will, can, dose, help, time, eat, need, use, life, one, well, hour, good, know, thing, morning, anything, people, levothyroxine	300 (13.67)
Treatment initiation	feel, day, like, get, just, better, week, start, much, medicine, want, sleep, make, pill, first, night, know, think, time, keep	343 (15.63)
Dose adjustment	medicaid, year, mcg, effect, side, dosage, increase, change, noticeable, differ, diagnose, hypothyroidism, hope, little, seem, removal, two, disease, old, great	396 (18.05)
Symptoms of pain	med, drug, problem, pain, even, stop, heart, low, cause, severe, bodies, bad, fatigue, headaches, anyone, worse, muscle, symptom, leg, look	379 (17.27)
Generic substitutability	synthroid, doctor, back, level, blood, test, put, normal, generic, never, trial, went, right, TSH, symptom, felt, told, year, said, brand	340 (15.50)
Appearance	weight, gain, month, hair, still, energized, tired, loss, start, see, lbs, time, lost, depressant, lose, always, skin, pound, dried, fall	436 (19.87)

[a] Total number of primary themes represented in each review. This is greater than the total number of reviews (1768) because some reviews had more than one theme tied for the primary concern.

Figure 1. Primary medication concerns unique to patients of different backgrounds.

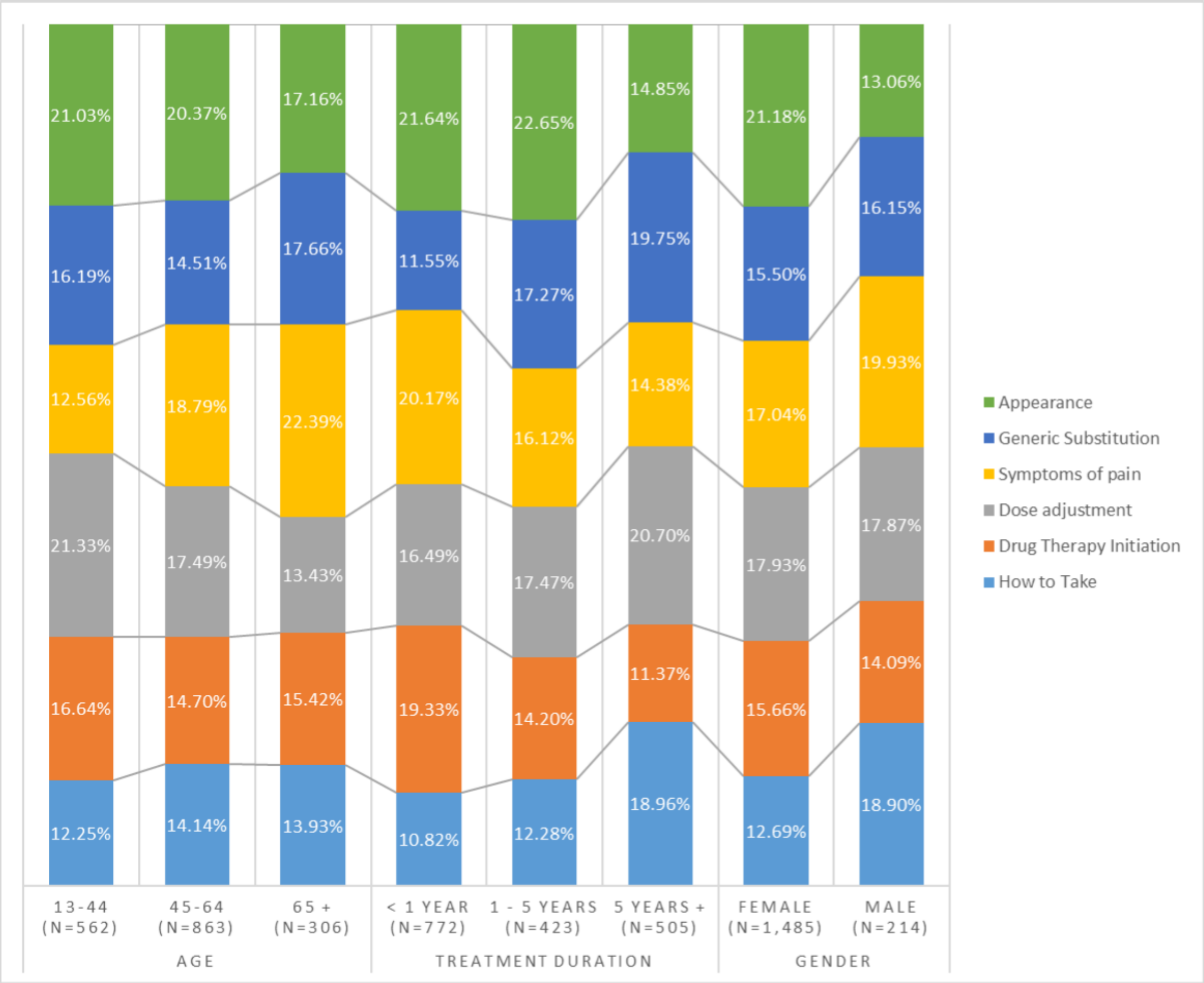


Table 3. Multiple regression analysis on factors affecting satisfaction score.

Characteristics	Coefficient (β)	Standard error	*P* value
Intercept	3.11	0.12	<.001
Primary medication concern			
Appearance (reference)	—	—	—
How to take the drug	0.68	0.12	<.001
Treatment initiation	–0.03	0.11	.81
Dose adjustment	0.44	0.11	<.001
Symptoms of pain	–0.29	0.11	.01
Generic substitutability	0.11	0.11	.32
Treatment duration, years			
More than 5 (reference)	—	—	—
Less than 1	–0.76	0.08	<.001
1-5	–0.38	0.10	<.001
Age, years			
45-64 (reference)	—	—	—
13-44	0.12	0.08	.14
65 and over	0.03	0.10	.76
Gender			
Female (reference)	—	—	—
Male	0.24	0.11	.03
Rating year	–0.04	0.02	.01

Dependency of Treatment Satisfaction on Primary Medication Concerns

The last aim of this study was to determine whether patient treatment satisfaction with THRT depends on what primary medication concerns they have, controlling for age, gender, and treatment duration. For this aim, each review's star rating of levothyroxine was treated as an interval scale rating. The 6 themes of patient medication concerns were dichotomized to indicate whether each medication concern was primary for the patient. As compared to the primary medication concern regarding appearance, the primary medication concern regarding how to take the drug resulted in a significantly better treatment satisfaction, up to as much as 0.68, with gender, age, treatment duration, and rating year being constant (Table 3). In other words, when patients had how to take the drug as their primary medication concern, their treatment satisfaction with THRT was 0.68 points higher out of 5 stars compared to those patients who had appearance as their primary medication concern. Treatment satisfaction also significantly increased (by 0.44 points) for the primary medication concern dose adjustment but reduced by 0.29 points (*P*=.01) for the primary medication concern symptoms of pain compared to the reference primary medication concern of appearance. Interestingly, treatment initiation and generic substitution did not have a significant impact on the treatment satisfaction with THRT relative to appearance.

Regarding the potential confounders controlled in the regression analysis, age did not result in any significant explanation of treatment satisfaction with THRT. For gender, male patients reported a significantly higher (β=0.24, *P*=.03) treatment satisfaction than female patients. For treatment duration, the longer the treatment duration, the higher the treatment satisfaction. In other words, as compared to the reviews written by patients who had taken levothyroxine for more than 5 years, the reviews written by patients who had taken the drug for shorter durations had significantly lower levels of treatment satisfaction.

Finally, multiple regression had the variable of rating year as an explanatory variable. The rating year represents the year the review was posted. For reviews posted in more recent years, the treatment satisfaction was slightly but significantly lower than the previous years. A 1-year increase in the rating year was associated with a decline of 0.04 points (*P*=.01) in the treatment satisfaction.

Discussion

Principal Findings

Above all, the NLP of patient medication reviews posted on a social network identified 6 distinctive themes of patient medication concerns on THRT: how to take the drug, treatment initiation, dose adjustment, symptoms of pain, generic substitutability, and appearance. Those 6 themes seem to cover all areas of medication needs of patients on THRT. The patient medication concerns on how to take the drug and dose adjustment quite well reflect the tough challenges associated with the task of taking levothyroxine appropriately. Levothyroxine requires scrupulous planning regarding when to take it. Many factors such as food, medications, and health conditions hamper the ability to maintain a desired range of thyroxine in the bloodstream [20]. It is critical to maintain the desired range because of the narrow therapeutic index and low absorption rate [21,22]. This study successfully identified those challenges into primary medication concerns how to take the drug and dose adjustment.

Likewise, the themes of symptoms of pain and appearance are well documented in the literature. Patients on THRT are known to complain about various symptoms of pain such as fatigue; loss of sleep; heart issues; headache; and muscle, leg, and body pain [2-4]. For this reason, a patient-reported outcomes instrument, Thyroid-Dependent Quality of Life, was developed to measure those symptoms of pain using several items on bodily discomfort [23]. Further, the most frequently mentioned medication concern was on appearance. A study reports that patients on THRT are often dissatisfied with levothyroxine treatment because of weight gain, hair loss, and dry skin [24].

Last, patients were quite frequently (15.5%) concerned about generic substitutability. Considering the price difference between brands and generic versions, patients would certainly rather use a generic version. To help ease patient concerns on generic substitutability, in 1997 the US Food and Drug Administration made the equivalence criteria stricter, from a range of 90% to 110% before to 95% to 105% after [25-27]. However, the issue of generic substitutability is still controversial. The American Thyroid Association reports that in a small fraction of patients higher rates of adverse events were associated with changes in levothyroxine preparations [6]. Furthermore, a reformulation of Merck's Lévothyrox with different excipients to reduce variability in dosing from 90% to 110% to 95% to 105% led to a huge controversy in France in 2017 (thousands of patients made complaints about adverse effects after taking the reformulated version [28]).

Documenting the uniqueness of primary medication concerns among different patients would emphasize the importance of tailored medication counseling. This study found that primary medication concerns were unique to patients of different genders, ages, and treatment durations. Female patients had appearance as their primary medication concern while male patients mentioned how to take the drug quite often. While it is apparent that females are quite attentive to appearance, it raises the question why males are more concerned about how to take the drug. Taking levothyroxine correctly requires careful planning regarding foods and time of administration. Perhaps females are accustomed to planning these things for family members, and thus, are more confident about taking the medication correctly than males.

This study also found that the youngest patients (aged 13 to 44 years) were very concerned about dose adjustment (21.33%) and appearance (21.03%) while the oldest (aged 65 years and older) were most concerned about symptoms of pain (22.39%). It is understandable that younger people were concerned about appearance. However, it requires some explanation why younger people were concerned about dose adjustment. Younger people must have just begun the levothyroxine treatment, and thus it would take longer for them to get the dose adjusted. On the other hand, older people were most concerned about the symptoms of pain because they were already suffering from many symptoms of pain resulting from aging. Patients with hypothyroidism would certainly experience symptoms of hypothyroidism unique to their age [22].

As the treatment duration became longer, patients were less concerned about appearance and treatment initiation and became more concerned about how to take the drug, dose adjustment, and generic substitution. Patients who had taken the drug for quite a while may have realized that it was more challenging to control hypothyroidism with levothyroxine. Once they realized that it is challenging, they would be more concerned about how to take the drug and dose adjustment. They would also be concerned about generic substitutability because a longer treatment duration makes it economically burdensome to take the expensive brand name drug.

Last, patient satisfaction with levothyroxine treatment depended on their primary medication concern. Patients who had the primary medication concerns of how to take the drug and dose adjustment were more satisfied with the levothyroxine medication therapy (up 0.68 and 0.44, respectively) compared to those whose primary medication concern was appearance. Evidently, patients who were more concerned about how to get the THRT right would have achieved better treatment outcomes. On the other hand, patients who had a primary medication concern of symptoms of pain were less satisfied with THRT (down 0.29 points) than the referents. Naturally, symptoms of pain would have negatively affected treatment satisfaction.

One of the interesting study findings was that the longer the treatment duration, the higher the treatment satisfaction. Patients who had taken the drug longer were more likely to have known how to take the drug correctly with the right dose and thus to have achieved better treatment outcomes. Moreover, males were more satisfied with the treatment than females. Because males were more concerned about how to take the drug but less concerned about appearance than females, they were more likely to achieve treatment success, which leads to higher satisfaction scores.

Practice Implication

Patients with hypothyroidism had unique primary medication concerns depending on their identity. In order to improve their treatment outcomes, it is necessary to provide tailored medication counseling, specifically focusing on their primary medication concerns such as how to take the drug, dose adjustment, symptoms of pain, and appearance.

Limitations

First, the LDA algorithm used in this research is a method of automatically discovering topics or themes by using statistical principles. Therefore, although LDA is advantageous in extracting hidden themes, the scientific qualities of the themes should be further validated. Furthermore, the researchers' subjectivity might have played an important role in the process of extracting hidden themes. For example, this study set the number of themes to be extracted to 6 based on its manageability. However, research has yet to come up with an easy way to choose an optimal number of themes [29]. Second, there may be a bias that arises from using online reviews from social media. Considering that people are more likely to report negative views than positive ones, the patient reviews used for this study may have come from patients dissatisfied with levothyroxine. Although negative views could better reflect patient concerns, they do not represent all views on levothyroxine. It is also important to note that this study used patient reviews from only one social media platform. Data collected from multiple social media platforms would have increased representativeness. Finally, the reported themes of patient medication concerns on levothyroxine treatment may not be valid because they were based on unstructured text data directly reported from patients. While those text data have their own merits, the reported themes need to be tested for their validity in future studies.

Conclusion

NLP of the patient reviews of levothyroxine posted on a social network produced 6 distinctive themes of patient medication concerns that covered all areas of levothyroxine treatment. Patients had unique primary medication concerns specific to their gender, age, and treatment duration. Furthermore, patients' treatment satisfaction on levothyroxine varied depending on what primary medication conditions individual patients had. These results inform the design of patient-tailored medication counseling to accommodate unique primary medication concerns of individual patients [30].

Acknowledgments

We thank the Research Institute of Pharmaceutical Science, College of Pharmacy, Seoul National University for supporting this study.

Conflicts of Interest

None Declared.

Multimedia Appendix 1

Word clouds of primary medication concerns.

[PNG File, 1MB - jmir_v20i10e11085_fig1.png]

References

1. IQVIA Institute for Human Data Science. 2018 Apr 19. Medicines use and spending in the US: a review of 2017 and outlook to 2022 URL: https://bit.ly/2yANWMX [WebCite Cache ID 72gHXAKPi]
2. Winther KH, Cramon P, Watt T, Bjorner JB, Ekholm O, Feldt-Rasmussen U, et al. Disease-specific as well as generic quality of life is widely impacted in autoimmune hypothyroidism and improves during the first six months of levothyroxine therapy. PLoS One 2016;11(6):e0156925 [FREE Full text] [doi: 10.1371/journal.pone.0156925] [Medline: 27257805]
3. Saravanan P, Chau W, Roberts N, Vedhara K, Greenwood R, Dayan CM. Psychological well-being in patients on 'adequate' doses of l-thyroxine: results of a large, controlled community-based questionnaire study. Clin Endocrinol (Oxf) 2002 Nov;57(5):577-585. [Medline: 12390330]
4. Wekking EM, Appelhof BC, Fliers E, Schene AH, Huyser J, Tijssen JGP, et al. Cognitive functioning and well-being in euthyroid patients on thyroxine replacement therapy for primary hypothyroidism. Eur J Endocrinol 2005 Dec;153(6):747-753. [doi: 10.1530/eje.1.02025] [Medline: 16322379]
5. Wiersinga WM, Duntas L, Fadeyev V, Nygaard B, Vanderpump MPJ. 2012 ETA guidelines: the use of L-T4 + L-T3 in the treatment of hypothyroidism. Eur Thyroid J 2012 Jul;1(2):55-71 [FREE Full text] [doi: 10.1159/000339444] [Medline: 24782999]
6. The ATA unveils new data to FDA on bioequivalence of levothyroxine. URL: https://www.thyroid.org/the-ata-unveils-new-data-to-fda-on-bioequivalence-of-levothyroxine/ [accessed 2018-07-17] [WebCite Cache ID 70yBTMWma]
7. Thyroid hormone treatment. URL: https://www.thyroid.org/thyroid-hormone-treatment/ [accessed 2018-07-17] [WebCite Cache ID 70y2i1MNk]
8. Dayan C, Panicker V. Management of hypothyroidism with combination thyroxine (T4) and triiodothyronine (T3) hormone replacement in clinical practice: a review of suggested guidance. Thyroid Res 2018;11:1 [FREE Full text] [doi: 10.1186/s13044-018-0045-x] [Medline: 29375671]
9. Doyle C, Lennox L, Bell D. A systematic review of evidence on the links between patient experience and clinical safety and effectiveness. BMJ Open 2013;3(1):1 [FREE Full text] [doi: 10.1136/bmjopen-2012-001570] [Medline: 23293244]
10. Manary MP, Boulding W, Staelin R, Glickman SW. The patient experience and health outcomes. N Engl J Med 2013 Jan 17;368(3):201-203. [doi: 10.1056/NEJMp1211775] [Medline: 23268647]

11. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Use of sentiment analysis for capturing patient experience from free-text comments posted online. J Med Internet Res 2013;15(11):e239 [FREE Full text] [doi: 10.2196/jmir.2721] [Medline: 24184993]
12. Fox S. The social life of health information. Washington: Pew Internet and American Life Project; 2011. URL: http://www.pewinternet.org/2011/05/12/the-social-life-of-health-information-2011/ [accessed 2018-08-09] [WebCite Cache ID 71XOmvYcf]
13. 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]
14. Yang C, Jiang L, Yang H. Tang X: Detecting signals of adverse drug reactions from health consumer contributed content in social media. 2012 Presented at: Proceedings of ACM SIGKDD Workshop on Health Informatics; 2012; Beijing.
15. Sadah SA, Shahbazi M, Wiley MT, Hristidis V. Demographic-based content analysis of Web-based health-related social media. J Med Internet Res 2016 Jun 13;18(6):e148 [FREE Full text] [doi: 10.2196/jmir.5327] [Medline: 27296242]
16. Salton G, McGill MJ. Introduction to Modern Information Retrieval. New York: McGraw-Hill, Inc; 1986.
17. Hornik K, Grün B. Topicmodels: An R package for fitting topic models. J Stat Softw 2011;40(13):1-30.
18. Zhan Y, Liu R, Li Q, Leischow SJ, Zeng DD. Identifying topics for e-cigarette user-generated contents: a case study from multiple social media platforms. J Med Internet Res 2017 Jan 20;19(1):e24 [FREE Full text] [doi: 10.2196/jmir.5780] [Medline: 28108428]
19. Choi HS, Lee WS, Sohn S. Analyzing research trends in personal information privacy using topic modeling. Comput Sec 2017 Jun;67:244-253. [doi: 10.1016/j.cose.2017.03.007]
20. Bach-Huynh T, Nayak B, Loh J, Soldin S, Jonklaas J. Timing of levothyroxine administration affects serum thyrotropin concentration. J Clin Endocrinol Metab 2009 Oct;94(10):3905-3912 [FREE Full text] [doi: 10.1210/jc.2009-0860] [Medline: 19584184]
21. Gottwald-Hostalek U, Uhl W, Wolna P, Kahaly GJ. New levothyroxine formulation meeting 95-105% specification over the whole shelf-life: results from two pharmacokinetic trials. Curr Med Res Opin 2017 Dec;33(2):169-174. [doi: 10.1080/03007995.2016.1246434] [Medline: 27718637]
22. Park KH, Lee EJ. Recent review on medical treatment of thyroid disease. J Korean Med Assoc 2012;55(12):1207. [doi: 10.5124/jkma.2012.55.12.1207]
23. McMillan CV, Bradley C, Woodcock A, Razvi S, Weaver JU. Design of new questionnaires to measure quality of life and treatment satisfaction in hypothyroidism. Thyroid 2004 Nov;14(11):916-925. [doi: 10.1089/thy.2004.14.916] [Medline: 15671770]
24. Levoxyl Professional Information. URL: https://www.rxlist.com/levoxyl-side-effects-drug-center.htm#overview [accessed 2018-07-17] [WebCite Cache ID 70yBa2KZY]
25. Brent G, Koenig R. Thyroid and anti-thyroid drugs. In: Goodman & Gilman's The Pharmacological Basis of Therapeutics, 12th Edition. New York: McGraw-Hill; 2011.
26. US Food and Drug Administration. 2018. Orange Book: Approved Drug Products with Therapeutic Equivalence Evaluations. 38th Edition URL: https://www.fda.gov/downloads/drugs/developmentapprovalprocess/ucm071436.pdf [WebCite Cache ID 71XQe8Yoe]
27. Darrel H. Current challenges in the management of hypothyroidism. US Pharmacist 2012 Jan:113-1678.
28. Casassus B. Risks of reformulation: French patients complain after Merck modifies levothyroxine pills. BMJ 2018 Dec 16;360:k714. [Medline: 29453193]
29. Zhao W, Chen JJ, Perkins R, Liu Z, Ge W, Ding Y, et al. A heuristic approach to determine an appropriate number of topics in topic modeling. BMC Bioinformatics 2015;16 Suppl 13:S8-S13 [FREE Full text] [doi: 10.1186/1471-2105-16-S13-S8] [Medline: 26424364]
30. Kalra S, Agarwal N, Aggarwal R, Agarwal S, Bajaj S, Bantwal G, et al. Patient-centered management of hypothyroidism. Indian J Endocrinol Metab 2017;21(3):475-477 [FREE Full text] [doi: 10.4103/ijem.IJEM_93_17] [Medline: 28553608]

Abbreviations

THRT: thyroid hormone replacement therapy
NLP: natural language processing
LDA: latent Dirichlet allocation

Edited by G Eysenbach; submitted 18.05.18; peer-reviewed by J Taylor, TH Collet; comments to author 21.06.18; revised version received 17.07.18; accepted 10.08.18; published 24.10.18.

Please cite as:
Park SH, Hong SH
Identification of Primary Medication Concerns Regarding Thyroid Hormone Replacement Therapy From Online Patient Medication Reviews: Text Mining of Social Network Data
J Med Internet Res 2018;20(10):e11085
URL: http://www.jmir.org/2018/10/e11085/
doi:10.2196/11085
PMID:30355555

Original Paper

Older People Negotiating Independence and Safety in Everyday Life Using Technology: Qualitative Study

Randi Stokke[1], RN, MEd

[1] Department of Health Sciences, Norwegian University of Science and Technology, Gjøvik, Norway

Corresponding Author:
Randi Stokke, RN, MEd
Department of Health Sciences
Norwegian University of Science and Technology
Postbox 191
Gjøvik, 2802
Norway
Phone: 47 951 58 088
Email: randi.stokke@ntnu.no

Abstract

Background: Due to demographic changes with an aging population, there is a demand for technology innovations in care services. However, technology innovations have proven difficult to implement in regular use. To understand the complexity of technology innovations in care practices, we need a knowledge base of the complex and diverse experiences of people interacting with established technologies.

Objective: This paper addresses the research gap in relation to understanding the microcontext of co-production of care involving established technologies integrated into care practices. The paper also aims to provide a framework for exploring what really happens when different actors use technology in care practices.

Methods: Participant observations and 22 interviews with actors using social alarms were conducted employing the critical incident technique. A stepwise deductive-inductive analysis was then performed.

Results: The results reveal how co-production of care assumes different meanings according to how actors use the technology. The results also show how technology innovation changes the dynamics between the actors and rearranges care practices. Independent and safe living is co-produced through performing bricolages and optimizing practice. Additionally, this opens up for unexpected results and bricolages as an integrated part of technology innovations.

Conclusions: This study illustrates how care services are always co-produced between the actors involved. By using aspects from science and technology studies, this paper provides a framework for exploring technology in use in care practices. The framework provides tools to unpack and articulate the process of co-producing services.

(J Med Internet Res 2018;20(10):e10054) doi:10.2196/10054

KEYWORDS

welfare technology; patient care; the social alarm; caring practices; home care services; science and technology studies; public service innovation; co-production

Introduction

Technology Innovations in Caring Practices

There is a persistent demand in public policy in western societies for increased technology innovation in community health care to meet challenges in the services caused by the “silver tsunami” and to facilitate active aging and independent living [1,2].

Technologies intended for care practices are often advocated as plug and play solutions to the challenges of facilitating safety and independent living and to avoid or postpone nursing home admission [3]. However, there is a discrepancy between these expectations and the complex reality in which these technologies are embedded [4-6]. The integration of technologies has proven difficult, and experiences so far have shown that it is very difficult to progress from pilot projects to regular use and then to scale up to other contexts [4,7-9].

There are many studies describing the effects of technology innovations in care practices [10] that attempt to identify drivers and barriers. However, such research seldom captures what happens after the pilot phase [11]. As the knowledge base is rather sparse, we need a more comprehensive understanding of how technologies integrated into health care practice work [10]. Moreover, we need a greater understanding of the practice of using established and integrated technology in regular service. My paper explores this by studying the use of the social alarm, an established technology that aims to provide safety and independent living for older people living at home. Focusing on this well-integrated and widely used technology allows us to study the emergence of personal, professional, and organizational issues that are little seen in new technology innovations. This is done by examining how people involved in using this technology co-produce safety and independent living within care practices and how they interact with the technology involved.

To address the research gap in understanding the complex work of established technologies in care practices, this paper positions itself within the newer public service innovation tradition that offers ways of understanding how public services are always co-produced by the actors involved in the service organization as an inevitable part of the service [12,13]. There has been little focus on the different actors' role, interaction, and co-production in a microcontext within the service innovation literature [13], and this paper contributes theoretically by suggesting a framework for exploring co-production in a microcontext, using constructs from science and technology studies (STS). In addition, it provides tools to unpack the process of co-production and address how this contributes to safety and independent living for older people with the social alarm.

The Social Alarm

Despite the major focus on technology innovations in care practices, social alarms still form the bulk of technologies in use in care practices and are widely in use in western societies [4,14]. The aim of a social alarm is to contribute to safety and independent living for frail older people. Even though well established and used since the late 70s, a previous review of social alarms illuminates diverse experiences and issues related to their use [15].

A social alarm consists of a unit placed centrally in the home and a pendant or wrist-worn device the end user can press when in need, as shown in Figure 1, to summon help from a dedicated responder.

Many actors are involved in the use of the alarm. The typical end user is an older dependent person living alone. Relatives, neighbors, care workers, and technology facilitators are all actors with different expectations, experiences, roles, and relationships with each other and the alarm [15,16].

Homecare workers usually offer the service needed when the social alarm is activated. As it is impossible to predict when it will be activated, the care worker might be anywhere in the homecare district at the time. The response time, therefore, varies from immediate response to several hours, depending on the number of care workers, their geographical location, and whether they are assisting other patients at the time.

The daily care practice is hard to plan, as care is unpredictable and emergent activities within a network of actors forming complex interactions. This demands flexibility and the ability to adjust plans, prioritizing some work and deprioritizing other work [17].

Figure 1. Illustration of the social alarm.

Theoretical Framework

This paper will draw on aspects of service innovation studies and STS to provide a fruitful contribution by exploring how the different actors involved co-produce safety and independent living for the end user of the social alarm. Innovation studies, including service innovation studies, and STS are two major adjacent fields inspired by many of the same theories. Both are concerned with technology innovations, social context, and the use of knowledge. Although thematically related, they are distinct research areas with limited interaction between them [18-21]. I will use the notion of script, domestication, and heuristics from STS as tools for framing, exploring, and understanding the co-production between actors involved. I will thereby add more pieces to the puzzle of understanding the complexity of technology innovation by articulating a framework for exploring the co-production between the actors and by grounding this framework empirically.

The concept of co-production is used in different ways, and its many definitions have been criticized for being rather blurred and lacking an empirical evidence base [22]. Co-production describes the relationships between different actors, and public service innovation literature states that services are always co-produced [23]. A service organization can suggest a service, but it is in the interaction between the service provider and the service user that the actual service is co-produced. Using the concept of co-production changes the traditional way of understanding a service as provided by the care service and received by the users of the service.

People are relational entities and co-produce care through “a relational, situated and embodied achievement in which people explore the right thing to do for themselves and their relationships” [24].

A co-production of practice (eg, a practice of using a technology) can provide different value for the different actors involved, such as safety and independence for the end user. However, the co-production might also lead to destruction of value [25,26] and in the worst case, failure of the technology integration.

Care research has shown that technology in use is always interwoven in complex networks of care practices that contribute to changing, shaping, and revealing new meanings of care in unforeseen ways [3-5,10,27,28].

Accommodating an integration of technology and co-producing care for the users implies a change in the care practices. This involves new interactions and changing roles and practices, redefining how actors live, work, and even identify their lives [10,28]. Responsibility is delegated to care workers and users [29,30], redefining their role from passive recipients to active participants and requiring them to become competent users of the technology [28]. This co-production involving changes in roles and responsibilities has been largely disregarded in previous research in the field [15] and will be addressed in this paper.

The co-production is a way of empowering as well as exploiting actors [31]. The experience of value is highly contextual and directly influenced by the expectations toward a service [25]. What goes on in the process of co-production is a complex and indistinct process that we need theoretical tools to explore.

Since a user’s perception is affected by both expectations and experience with the service, the moment these two collide is the moment of truth. This moment is, in fact, a continuum, a process from expectations to experience [12,25].

The notion of script and domestication offers us tools for exploring and analyzing the practical work of co-production of care and underlines how using a social alarm can lead to values such as safety and independence for the end user.

The script metaphor describes technologists’ vision of a technology’s function, the expectations of how the users will relate to the technology and vice versa [32-35]. Social alarms can be scripted in different dimensions: as technological devices, as integrated services, and how values come into play for the actors involved [36].

The domestication metaphor, inspired by the process of domestication of wild animals, provides a way of exploring this continuum in depth and showing the relational process of how an object gradually becomes a part of everyday living [37,38]. This involves back and forth battles of values, pride, resistance, refusal, and tension in the interaction between the technology and humans involved [33,38].

From the domestication process, a more or less stable relationship is established between the actors and the technology. Pols [10] describes this as taming and unleashing in care practices where 4 heuristics emerge. Following these relationships enables us to explore the interactions and co-production in depth.

There are examples where human-technology interaction works exactly as planned/scripted. Furthermore, sometimes the actors reject the technology altogether—there is no domestication, which causes the failure of technology projects. However, this paper focuses on technology in use. Technology in use in this context means that the technology is integrated and adopted into a care practice.

As technology innovations become an established part of practice, the focus fades and the integration process becomes more tacit [39]. You might say that the technology tends to move off the radar as it becomes domesticated. However, even if domesticated, a technology is never completely fixed. The technology as an artifact opens up for different interpretations of how to use it, what to think of it, and what feelings it inspires, etc. [40,41].

The practice of descripting and domestication described above often involve small ad hoc innovations called bricolages. These are simple, unplanned solutions using resources at hand as an answer to a problem [4,42]. Table 1, an expanded version of Pols [10], illustrates different heuristics that play out in practice involving technology in use and how this might lead to bricolage. Thus, Table 1 provides an illustration of how both the technology and human actors are active parts of the co-production of care practices involving technology.

Table 1. Describing heuristics of taming and unleashing technology and actors. An expanded version of Pols (2017).

Heuristics	Description of the heuristics
The techology is used in accordance with the scripts	Actors adopt, integrate, and domesticate the technology in accordance with expectations, and the technology is part of co-producing care practices.
Taming the users	Technologies sometimes tame users by making them dependent on the technology and making them adjust their lifes according to the technology.
Unleashing the users	Technologies can unleash users, making them request new services from the technology. Here the script and intention of the technology are not meeting the demands and expectations from the actors involved, thus leading to bricolage or dissatisfaction with the technology.
Taming the technology and unleashing practices	Actors tame the technologies by using them to pursue their goals, either by exploiting only some possibilities the technology offers or by finding new ways of use, often through bricolages, often in other ways than scripted and intended by designers and vendors. Sometimes the technologies unleash unexpected and completely new areas of use.
Nonuse	The users reject the technology altogether; domestication does not occur.

Acts of bricolage can add up to significant changes in routines or use of technology but they are often hidden and unarticulated in the co-production of daily practices [43]. Nurses often perform bricolage in care work to solve problems on the spot [44].

According to Star and Strauss [45] and Allen [17], care work largely incorporates this way of working where actors need to adjust the practice or the technology [4,17]. Bricolage often allows users and family members to take the initiative in co-producing care by finding better solutions [46]. By collecting bricolages and bringing them to the attention of colleagues and management, it is possible to turn them into useful innovations to improve the quality of the service [42].

Framing Co-Production of Technology in Use

Bridging the different constructs described in an integrated framework contributes theoretically by offering ways of exploring how the co-production practice in care work involving technology might be explored, as presented in Figure 2.

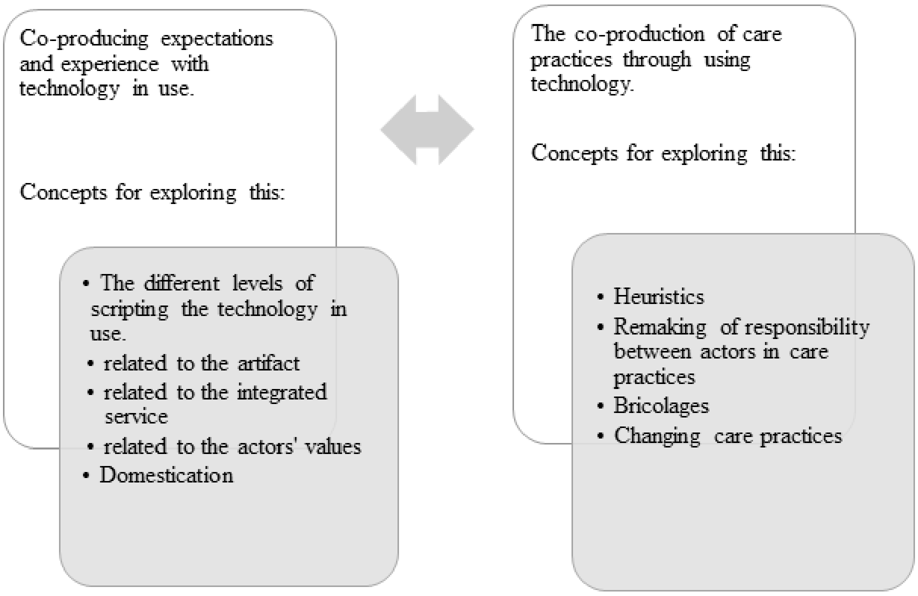
This constitutes a continuum of tools for exploring and understanding the complex reality of co-produced care practices with technology in use. The model's linearity is for analytical purposes, as the lived reality is dynamic, with processes going back and forth.

This framework provides tools for understanding the use of established technologies in care practices by adding STS concepts to explore the co-production that always takes place between the different actors involved. Later in the paper, this will be related to the data from the empirical material related to the social alarm.

The aim of the paper was to explore and interpret different actors' experience with an established technology innovation in care practices and discover how actors use and interact with social alarms and what strategies they apply when co-producing safety and independent living. The study focuses on the experience of safety and independence, as these are expressed aims for the social alarm.

The research question examined: How do older people pursue, maintain, and negotiate independence and safety in everyday life by using social alarms?

Figure 2. Integrated framework.



Methods

Context, Settings, and Sample

The study was conducted within a homecare service in Norway in the period from August 2014 to December 2015, combining 1-week full-time participant observation in 2 municipalities and in-depth interviews (n=22) with actors involved in the use of social alarms. This provided a comprehensive picture of the respondents' experiences with social alarms, promoting an in-depth understanding of this technology in use.

In Norway, all inhabitants are entitled to care services at home if required [2,47,48]. Two municipalities were strategically chosen for maximum diversity, representing both typical and diverse local communities in Norway as displayed in Table 2.

One municipal homecare manager in each municipality recruited participants for interviews. Older people, relatives, and care workers were selected to be interviewed according to the inclusion criteria described in Table 3.

A strategic participant selection was conducted aiming for maximum diversity. A description of included respondents is presented in Multimedia Appendix 1. The end users lived either in their own home or in care flats. The latter are small flats, often organized in proximity to a nursing home. The fieldwork involved participant observation in homecare services in both municipalities in the various cultural contexts. This entailed accompanying care personnel responsible for receiving alarm calls on their shifts while visiting end users.

The interviews were conducted with people possessing a social alarm (n=11), relatives of people possessing the alarm (n=4), care workers with different backgrounds and responsibilities (n=6), and key workers and managers (n=3). Some respondents had several roles. The 22 interviews were conducted in the respondents' homes or at home care service facilities and lasted between 30 minutes and 2 hours.

In this study, the respondents told hundreds of stories, providing extensive material including the narratives presented later. All interviews were audio recorded and fully transcribed verbatim. NVIVO 11 (QSR International) software for data analysis was used as a tool when organizing, analyzing, and finding insights into the material.

Interview Design

Critical incidence technique (CIT) methodology was applied, and the respondents were asked to describe incidents related to the social alarm. CIT is a practically oriented, commonly used explorative approach that facilitates insights into the complexities of an event and the interactions between actors involved [49,50]. This provided rich stories of critical incidents balanced with normal use descriptions. The method is attentive to the way practice is lived, generates rich material, and uncovers tacit understandings of an incident, including affective, cognitive, and behavioral elements [51,52]. Critical incidents can never be seen as isolated but rather as integrated into contexts. Narratives of lived experiences thereby emerged as presented in the results section. CIT is used within different research traditions, and this study is aligned with a phenomenological-interpretivist tradition as developed by Elizabeth Chell. Incidents are described as something emerging from the practice, embedded in the actor's perspective [50,51,53].

Ethics Approval and Consent to Participate

The Data Protection Official for Research in Norway granted approval for the project (project number 38605). All interview participants gave informed written consent. The municipal health care management gave written consent to collect data within the homecare services. All health care professionals received written and verbal information about the study and data collection and were informed that they could decline to be observed or interviewed. Prior to the observational study, the health care workers informed patients about the study and asked if researchers could observe the encounter.

Table 2. Characteristics of the municipalities included in the study.

Characteristics	Municipality 1	Municipality 2
Inhabitants	30,000	2600
Geography	Midsize city; inland	Rural district; northern coastal area
Responders	Homecare personnel	Call center that contacts homecare personnel when necessary

Table 3. Recruiting criteria for the interviews.

Respondents	Recruiting criteria
End users	• Possessed a social alarm for more than 1 year • Varying experiences with the alarm • Both sexes, a variety of ages, living conditions, and dependency
Next of kin	• Difference in relationships and living distance from the end user
Care workers	• Experience with responding to the social alarm • Varying professions and responsibility related to the alarm

Analysis

Detailed observational field notes and transcriptions from the interviews were analyzed thematically using a stepwise deductive-inductive analysis described by Tjora [54], who was inspired by Strauss and Corbin [55,56]. The analytical focus was on participants' perspectives and experiences in accordance with CIT [49].

The analysis started with an empirically close coding of the transcripts resulting in 68 codes focusing on participants' perspectives. This inductive coding related to the actor's experiences and co-production with the social alarm. This was further analyzed and grouped into 11 categories based on the actors' voices [54]. Three themes of relevance to this paper formed the empirical-analytical basis: attaching to the alarm, interacting with others, and tinkering and bricolage. These themes embody how the different actors involved co-produce safety and independent living for the end users. The themes apply to all respondents in the study in various ways.

By going back and forth between the theory and the themes, 3 narratives emerged from this analysis, using Figure 2 as a sensitizing tool. Purposeful sampling provided a way of studying these rich cases in depth [57], as the narratives are suited for illustrating the themes from different perspectives and provide rich descriptions of different user styles. This illustrates how the social alarm opens up for different co-production practices and contributes to different experiences of safety and independent living when integrated into daily lives.

Results

Co-Producing Safety and Independent Living With the Social Alarm as a Life Saver

The 3 anonymized narratives presented reflect different unique lived experiences related to the social alarm and how safety and independent living are co-produced between the different actors and the social alarm. This illuminates different aspects of the empirical material corresponding to the 3 themes that emerged from the analysis.

The respondents recounted dramatic and possibly life-threatening events. Even so, this first narrative particularly stood out as "Anna," her daughter "Turid," and several of the care workers related it unsolicitedly.

Anna is a 96-year-old widow. She lives alone in her house in a depopulated rural area. She and one other person still live in the hamlet. She got the social alarm 16 years ago after suffering a stroke. She describes a strong attachment to the alarm and says that she feels safe, trusting that she can get help if needed. She has 5 adult children, all living far away. However, they speak daily on the phone, and someone always comes home during the holidays. Anna suffers from several chronic conditions, making her dependent and frail. However, she manages on her own and walks with a stick. This story took place in late wintertime when it was very cold and there was a lot of snow on the ground.

> *I was going outside with a bag of old newspapers. Then I fell on the ice. Broke my hip. I was lying there completely immobilized. I was wearing the alarm at the time; otherwise, I would not be here today. It was a quarter to ten in the morning when I fell, and they would not have started looking for me before six o'clock in the afternoon. However, I had the alarm, so I pressed the pendant and got help from the homecare nurse and ambulance in just a few minutes. I was conscious the whole time; the nurse came immediately because she was nearby. She went inside for a pillow and a blanket, which she laid under me until the ambulance came and took me to the hospital. The alarm saved my life that time, that's for sure. And many other times as well, as things are. I would be dead by now for sure if I didn't have the alarm.*

Interacting with others was central for Anna's use of the social alarm, and the collaboration with the care worker made her feel safe and attended to.

When Anna talked about her attachment to the alarm, she said, "The most precious thing I have is the social alarm." She described how the alarm was crucial for her ability to feel safe and to be able to stay at home. She was not afraid for herself, she said. However, she did not want her children to worry.

Her daughter Turid said that she felt her mother was fairly safe, as long as she had the social alarm, as they had experienced that she got help when in need. When asked what the alarm means for her as next of kin she answered:

> *Very much. It really does. We do talk on the phone every day, but she has had a stroke you know. She is paralyzed on her right side. She cannot grip things with her right hand. And her right foot, she sways a bit, and overbalances easily. We are a bit insecure regarding her staying at home. She is clear-headed and wants to continue living at home. And as she says: I do have the alarm. Yes, but as I say, she has been lucky when falling, not falling on her left side. Then she would not be able to use the alarm. The last time she fell, I started to wonder how ethically right it is, we all feel that way.*

Turid describes a fragile dependent mother as borderline in terms of whether she is able to live alone with the social alarm as a safety net—they are managing, but only just.

Depending on the Social Alarm When Co-Producing Independence

Although many of the stories told in this study were dramatic, some are about undramatic but still decisive experiences. Some end users had never activated the social alarm due to emergencies; one of them is "Jon."

Jon is a man in his early sixties with a progressive neurological disease that has partly restricted him to a wheelchair. His health is deteriorating, but he still manages to live in his house, with visits from the care worker for weekly medication. He appears emotional regarding his attachment to the social alarm and repeatedly praises it for making living at home possible.

> *...the possibility of having one of these [enthusiastically waves the pendant], especially when*

> *you live alone and are still going down to the basement for firewood. Then it’s crucial, the social alarm, because it actually works... Hurray for the alarm!*

Jon has had a social alarm for some years now. He always has it in the pocket of a leather waistcoat with his mobile phone in the other pocket.

Jon describes himself as a very engaged, active man with a positive attitude. When Jon talks about his attachment to the alarm, he describes how it makes him feel safe and allows him to keep doing things without worrying about falling and not getting help. At the same time, he adds that he didn’t care for the alarm in the beginning. He regarded it as an indicator of his declining health and felt that by accepting the alarm, he was accepting deterioration. He needed time to get used to it before accepting it. This aligns with other studies describing how end users chose to use the alarm but did not care for it [36,58,59]. Jon had a brother with the same disease as himself. The two of them had different approaches to life with a disability, Jon said. While he really wants to manage on his own and stay active, his brother in contrast accepted his decline sooner and became more passive. One example is that Jon only uses his wheelchair when necessary, while his brother embraced the wheelchair and never got up again once he sat down.

Jon described how important it is for him to decide for himself and to be in a dialogue, interacting with the care service. He described meetings where care workers suggested that he moved to a care flat, but he wants to stay in his house. He also gets positive feedback from the care workers regarding his independence and humor and how he likes to make jokes.

Jon had never needed to activate the alarm due to an emergency, and he talked about how he sometimes felt insecure about whether it really worked and that he sometimes activates it just to check.

> *I succumbed to the temptation a few times and pressed the pendant. “Hello, this is me. If you are nearby, could you please come by? I have not been to the grocery store and need some warm food.” Then I started rattling out anything I could think I needed. They gradually started laughing. They know me, they’ve been here before and knew I’m hopeless that way. Black humor, you could say...*

It was important for Jon to be valued as independent and with a sense of humor. The quote describes how he managed to maintain his humorous attitude toward the care personnel while at the same time co-producing safety by testing the social alarm. Moreover, the care personnel allowed him to do so. The alarm has automatic functional testing so manual testing could be regarded as unnecessary. However, this co-production made him feel safe and confirmed that he could trust the alarm. He was dependent on the alarm to be safe enough to live independently, even though he had never needed to activate it.

Creative Co-Production With the Social Alarm

The last narrative is about “Peter” and “Marie.” Their daughter “Kari” told the story. Peter and Marie were an old married couple living in a care flat. People living there are usually dependent and frail, as were Peter and Marie, who moved from their house when their health was deteriorating. Peter tended to fall and he sometimes passed out due to a drop in blood pressure. Then his fingers got numb and he could not manage to press the pendant. Marie was the one wearing the alarm. She was physically vigorous, even though she had had a stroke some years back and suffered from dementia. Marie did not understand how she should call for help when her husband got ill, and Peter was too heavy for Marie to manage. Kari describes how her parents had a combined attachment to the social alarm.

> *When he experienced falls in blood pressure, he passed out for a while. Sometimes she pressed the pendant, sometimes not. She got confused when these things happened, but then he woke up and said, “Press the button!”*
>
> *After they moved to the care flat, he fell several times. Sometimes he passed out, but not always. Then he told her to press the pendant. This could be early in the morning, or if he needed to get up to pee at night, ... and then she pressed the pendant. I don’t think she did it before he told her to. No, but then they got help, and I believe that if they hadn’t had the alarm, then he might have been lying there for 2-3 hours, maybe more. There are many examples like this.*

Marie managed to press the alarm pendant when Peter instructed her. Together they managed to co-produce what neither of them managed alone by tinkering with the use of the alarm. Kari could not tell how this arrangement came about, but she thought it was done in collaboration between Peter, Marie, and the care workers.

Then Peter died. Marie still has the social alarm, but Kari does not believe her mother is capable of activating it when in need, as she has never activated it since her husband died. She is very dependent, and her dementia is progressing.

Discussion

Principal Findings

The key objectives of this paper were to explore and interpret how through interaction with the social alarm different actors co-produce safety and independent living for older people with the alarm. A further aim was to provide a framework for exploring co-production in care practices when technology is integrated.

The framework illustrated in Figure 2 provides theoretical tools to explore and interpret the co-production. This entails using the social alarm to pursue, maintain, and negotiate independence and safety for end users. Examples from the narratives presented in the paper illustrate different user styles and work and interaction with the social alarm in use as further presented.

Attaching to the Alarm

Figure 2 presents scripting and domestication as tools for exploring the co-production of expectations and experiences with technology in use. It further illustrates how a technology in use can be scripted in dimensions related to the artifact, service, and values that come into play. When using Figure 2 as a sensitizing concept, the analysis revealed that the respondents in this study co-produce attachment to the alarm. This relates to expectations described in the scripting and experiences through the domestication process. The advocated scripting of the social alarm describes a technology that enables safe and independent living. Most respondents in this study agreed enthusiastically with this script. Jon's story differentiates this attachment, as he did not care for the alarm at first, struggling with the less articulated script of it being suited for dependent and frail people. He interpreted it as a declaration of his deterioration. He gradually came to accept the alarm, now describing himself as dependent on the technology to stay independent. This illustrates the multiple scripting of the technology and how the relationship with the alarm might change over time through domestication. Exploring scripts and domestications provided insights into the co-production of expectations toward the social alarm and allowed us to articulate and explore the co-production of safety and independence in depth.

Interacting With Others

Service innovation literature states that a service is always co-produced, as illustrated in Figure 2. So, what does this imply for the actors in their everyday practice?

The social alarm only works through co-production between the actors, including the technology as an actor. It is tenable and fragile and needs a collective, mutual persuasion in order to work and make sense. The person possessing the alarm needs to wear the pendant and activate the alarm to get help. The care worker has to answer the alarm calls and effect proper response. The active participation requirement largely means that responsibility for the safety of end users is delegated to the end users themselves, care workers, and the technology. This demonstrates how the social alarm is not "strong" in the sense of a powerful stand-alone technology. It also illustrates how the social alarm changes the dynamics between actors and technology and rearranges care practices.

Table 1 provided us with examples of different heuristics that arise in the interaction between human actors and technology. If we integrate examples from the narratives into Table 1, we find that these represent different heuristics, as described in Table 4.

Using Figure 2 enables us to focus on the script and domestication and how humans and technologies co-produce safety and independence in different ways, as illustrated in Table 4. Interpretative flexibility and contextual factors contribute to these different heuristics, and different practices emerge as we follow the suggestion of Pols [10] to study technologies in the context and network they are integrated into.

Co-Production by Tinkering and Bricolage

The last column in Figure 2 provides us with a focus on how care is co-produced through the integration with the technology. Sometimes the technology does not quite fit the users' needs or even their ability to handle the technology, and there is a need to work around this, creating bricolages. Bricolage comes into play in the co-production between end users and other actors as described in Table 4.

Bricolage appears when technology fails or does not meet the needs of the user, as when Jon activates the alarm to check whether it works. Bricolages also occur when the users do not meet the demands of the technology as illustrated by the example of Peter and Marie and their co-production of activating the alarm.

Table 4. Different practices illustrated by examples from the narrative.

Heuristics	Examples from the narratives of the co-production between technology, actors, and care service to meet the actors' needs.	How actors involved co-produce care practices
Use in accordance with expectations	Anna and her family co-produce the interaction with the technology in accordance with the script, securing help when in need.	Co-producing safety and independence as expected by advocates of the social alarm and described in the scripts.
Taming the users	All narratives provide examples of how the end users are dependent on the technology but in very different ways. They are dependent on the social alarm to be able to live independently.	The users are able to co-produce the value of staying at home and feeling safe by interacting with the technology and other actors in the service.
Unleashing the users	Jon and the care workers co-produce a new service by allowing him to activate the alarm to check whether it works, even though testing is done automatically.	Using the alarm in an unpredictable way through bricolage. Co-producing safety and independence.
Taming the technology and unleashing practices	Peter and Marie are co-producing a way of taming the technology by co-producing ways of using the alarm. Both Marie and Peter's collaboration and Jon's workabouts are good examples of how the technology can unleash unexpected practices.	Co-production of independence and safety. By this, the time frame of use of the social alarm is extended through bricolage. If the municipality focuses on these unexpected co-production practices, they can use them for quality improvement and potential innovations in the service.
Nonuse	This was not relevant in this study as the research focused on the technology in use.	There is an important distinction between not activating the social alarm and not using the alarm, as we can see with Jon, who is dependent on the alarm although he has never activated it in an emergency.

Technology in care practices often has a time frame when it is possible for the users to use the technology, especially when the end users' health is deteriorating as is often the case with elderly and chronically ill people [36]. The co-production between Marie and Peter made it possible to prolong the period of use. Neither of them was able to use the social alarm individually, but in collaboration, they were able to co-produce safety.

It is evident that vulnerability and even luck are involved in the success of the social alarm in use. Anna's daughter describes how Anna is fortunate to have fallen on her right side. If she had fallen on her left side, she would not have been able to press the pendant due to paralysis. Luckily, the care worker was nearby when Anna broke her hip. If Peter had remained unconscious one day, Marie might not have been able to activate the alarm. It seems that sometimes there are narrow margins for the alarm to provide safety. This aligns with the argument of Procter and Greenhalgh [1] that although customization or bricolages might have advantages in the form of usability, they may also endanger patient safety.

Today these technologies are advocated as plug and play and one-size-fits-all user technology. This is an illusion. We should instead study the bricolage that emerges, as the stories illustrate, and make sure that quality is improved in a safe way for the end users. We would then be able to use more of the technology systematically in the long term and develop possible service innovations. One simple example is to copy Jon's use by encouraging end users to press the pendant occasionally, giving them experience in activating the alarm and reassurance that it works.

To ensure quality in care practices, we need to ground the integration of the technology and other parts of the care practices in the end users' lived experience [1].

Social alarms have been around for more than 30 years, and there are still ongoing processes of bricolage and co-production between actors to make them work. The technology is well established and thereby tends to be viewed as normalized and closed. However, when taking the actors and context in which the practice takes place into account, we find ongoing co-production processes where the users redefine and renegotiate the purpose and practical use of the alarm. According to Oudshoorn and Pinch [60], users will always find new ways of using familiar technologies. This bricolage can be seen as a kind of continuous service innovation, which seems to be a precondition for the integration of technology in everyday life.

Conclusions

This paper focuses on the co-production of established technology in use. Using concepts from STS in the framework as presented in Figure 2 facilitated a comprehensive approach to studying technology integrated into care practices. The framework provided a tool to unpack and articulate the process of co-producing safety and independent living, which are the aims of social alarms. The study shows how practice involving an established simple technology may actually comprise complex practices that we need to thoroughly explore to understand. This would provide us with a more comprehensive picture of the interaction between them and the complexity at hand. The framework proved relevant for exploring technology innovations in care practices empirically.

Even with successful practices where the care and use of technology are grounded in the end users' experience, bricolages are still seen. They should be regarded as possibilities for quality improvement in addition to tinkering with technology that does not work. Care work mainly consists of co-production between different actors in networks, with technologies as integrated components. Therefore, technology in use must be viewed as a collaborative activity.

This paper also reveals how the experience of the technology, as well as the possibility to use it, changes over time. Technology integrated into care practices can, therefore, not just be implemented and left alone to function; it must be adjusted and followed up as technology in use, constituting an everlasting process.

This study suggests a framework for exploring co-production when using technologies in care practices. However, this research is limited to social alarms, and several areas should be investigated further. This study focuses on examples where the technology mainly works and is successful. Empirical exploration of more complicated technologies going forward could promote better analytical clarity and contribute to the validation of the framework. Furthermore, studies of the changes in the different actors' roles in the co-production of care practices would provide valuable insights.

Acknowledgments

Sincere thanks to supervisors Professor Lars Fuglsang, Professor Rolf Rønning, Professor Ragnhild Hellesø, and Maren Sogstad, PhD. Thanks to all participants who took part in the study and the managers in homecare practices in both municipalities. The author is a doctoral candidate employed at the Norwegian University of Science and Technology. No external funding was received for this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Providing an overview of the included respondents.

[PDF File (Adobe PDF File), 40KB - jmir_v20i10e10054_app1.pdf]

References

1. Procter R, Greenhalgh T, Wherton J, Sugarhood P, Rouncefield M, Hinder S. The day-to-day co-production of ageing in place. Comput Support Coop Work 2014;23(3):245-267 [FREE Full text] [doi: 10.1007/s10606-014-9202-5] [Medline: 26321795]
2. Huber M, Rodrigues R, Hoffmann F, Gesior K, Marin B. Facts and Figures on Longterm Care for Older People: Europe and North America. Vienna: European Centre for Social Welfare Policy and Research; 2009.
3. Pols J, Willems D. Innovation and evaluation: taming and unleashing telecare technology. Sociol Health Illn 2011 Mar;33(3):484-498 [FREE Full text] [doi: 10.1111/j.1467-9566.2010.01293.x] [Medline: 21241338]
4. Procter R, Wherton J, Greenhalgh T, Sugarhood P, Rouncefield M, Hinder S. Telecare call centre work and ageing in place. Comput Supported Coop Work 2016 Jan 15;25(1):79-105. [doi: 10.1007/s10606-015-9242-5]
5. Sánchez-Criado T, López D, Roberts C, Domènech M. Installing telecare, installing users. Sci Technol Hum Val 2014 Jan 09;39(5):694-719. [doi: 10.1177/0162243913517011]
6. Essén A, Conrick M. New e-service development in the homecare sector: beyond implementing a radical technology. Int J Med Inform 2008 Oct;77(10):679-688. [doi: 10.1016/j.ijmedinf.2008.02.001] [Medline: 18514021]
7. Andreassen HK, Kjekshus LE, Tjora A. Survival of the project: a case study of ICT innovation in health care. Soc Sci Med 2015 May;132:62-69. [doi: 10.1016/j.socscimed.2015.03.016] [Medline: 25795426]
8. Halford S, Lotherington AT, Obstfelder A, Dyb K. Getting the whole picture? Inf Comm Soc 2010 Apr;13(3):442-465. [doi: 10.1080/13691180903095856]
9. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
10. Pols J. Good relations with technology: empirical ethics and aesthetics in care. Nurs Philos 2017 Jan;18(1):1. [doi: 10.1111/nup.12154] [Medline: 27758077]
11. Essén A, Lindblad S. Innovation as emergence in healthcare: unpacking change from within. Soc Sci Med 2013 Sep;93:203-211. [doi: 10.1016/j.socscimed.2012.08.035] [Medline: 23021848]
12. Osborne SP, Strokosch K. It takes two to tango? Understanding the co-production of public services by integrating the services management and public administration perspectives. Brit J Manage 2013 Aug 07;24:S31-S47. [doi: 10.1111/1467-8551.12010]
13. Alford J. Co-production, interdependence and publicness: extending public service-dominant logic. Public Manag Rev 2015 Nov 20;18(5):673-691. [doi: 10.1080/14719037.2015.1111659]
14. Heinbüchner B, Hautzinger M, Becker C, Pfeiffer K. Satisfaction and use of personal emergency response systems. Z Gerontol Geriatr 2010 Aug;43(4):219-223. [doi: 10.1007/s00391-010-0127-4] [Medline: 20814795]
15. Stokke R. The personal emergency response system as a technology innovation in primary health care services: an integrative review. J Med Internet Res 2016 Jul 14;18(7):e187 [FREE Full text] [doi: 10.2196/jmir.5727] [Medline: 27417422]
16. Mørk E. Statistisk sentralbyrå. 2015. [Kommunale helse- og omsorgstjenester 2014: statistikk om tjenester og tjenestemottakere] URL: https://www.ssb.no/helse/artikler-og-publikasjoner/_attachment/242516?_ts= [WebCite Cache ID 72qPSxE3I]
17. Allen D. The Invisible Work of Nurses: Hospitals, Organisation and Healthcare. London: Routledge; 2014.
18. Bhupatiraju S, Nomaler O, Triulzi G, Verspagen B. Knowledge flows: analyzing the core literature of innovation, entrepreneurship and science and technology studies. Res Policy 2012 Sep;41(7):1205-1218. [doi: 10.1016/j.respol.2012.03.011]
19. Fagerberg J, Verspagen B. Innovation studies: the emerging structure of a new scientific field. Res Policy 2009 Mar;38(2):218-233. [doi: 10.1016/j.respol.2008.12.006]
20. Hyysalo S, Jensen T. Changing Innovation Collectives and Involvement Strategies. Volume 42. London: Routledge; 2016.
21. Lundvall B. Innovation studies: a personal interpretation of “the state of the art”. In: Fagerberg J, Sloth Andersen E, Martin BR, editors. Innovation Studies: Evolution and Future Challenges. Oxford: Oxford University Press; 2013.
22. Nabatchi T, Sancino A, Sicilia M. Varieties of participation in public services: the who, when, and what of coproduction. Public Admin Rev 2017 Mar 30;77(5):766-776. [doi: 10.1111/puar.12765]
23. Brandsen T, Pestoff V. Co-production, the third sector and the delivery of public services. Pub Manag Rev 2006 Dec;8(4):493-501. [doi: 10.1080/14719030601022874]
24. Moser I, Thygesen H. Exploring possibilities in telecare for ageing societies. In: Barnes M, editor. Ethics of Care: Critical Advances in International Perspective. Bristol: Bristol Policy Press; 2015:111-123.
25. Osborne SP, Radnor Z, Strokosch K. Co-production and the co-creation of value in public services: a suitable case for treatment? Pub Manag Rev 2016 Mar 09;18(5):639-653. [doi: 10.1080/14719037.2015.1111927]
26. Skålén P, Engen M, Magnusson P, Bergkvist L, Karlsson J. Public service innovation: a public service dominant logic view. 2016 Presented at: Proceedings of the 26th Annual RESER Conference; Tiziana Russo-Spenaand Cristina Mele; 2016; Naples.

27. Pols J. Telecare: what patients care about. In: Mol A, editor. Care in Practice. Bielefeld: Transcript Verlag; 2010:171-194.
28. Oudshoorn N. Telecare Technologies and the Transformation of Healthcare. Basingstoke: Palgrave Macmillan; 2011.
29. Nicolini D. The work to make telemedicine work: a social and articulative view. Soc Sci Med 2006 Jun;62(11):2754-2767. [doi: 10.1016/j.socscimed.2005.11.001] [Medline: 16343724]
30. Oudshoorn N. How places matter: telecare technologies and the changing spatial dimensions of healthcare. Soc Stud Sci 2012 Feb;42(1):121-142. [doi: 10.1177/0306312711431817] [Medline: 22530385]
31. Essén A, Värlander SW, Liljedal KT. Co-production in chronic care: exploitation and empowerment. Eur J Marketing 2016 May 09;50(5/6):724-751. [doi: 10.1108/EJM-02-2015-0067]
32. Joyce K, Mamo L. Graying the cyborg. In: Calasanti T, editor. Age Matters: Realigning Feminist Thinking. New York: Routledge; 2006.
33. Berker T. Domesticating spaces. Space Cult 2011 Jul 07;14(3):259-268. [doi: 10.1177/1206331211412259]
34. Akrich M. The de-scription of technical objects. In: Bijker W, editor. Shaping Technology/Building Society: Studies in Sociotechnical Change. Cambridge: MIT Press; 1992:205-224.
35. Woolgar S. Configuring the user: the case of usability trials. The Soc Rev 1990;38(S1):58-99.
36. Stokke R. “Maybe we should talk about it anyway:” a qualitative study of understanding expectations and use of an established technology innovation in caring practices. BMC Health Serv Res 2017 Sep 15;17(1):657 [FREE Full text] [doi: 10.1186/s12913-017-2587-3] [Medline: 28915809]
37. Sørensen K. Domestication: the enactment of technology. In: Berker T, editor. Domestication of Media and Technology. Berkshire: Open University Press; 2007.
38. Silverstone R. Domesticating domestication: reflections on the life of a concept. In: Berker T, editor. Domestication of Media and Technology. Berkshire: Open University Press; 2007:229-247.
39. Bartunek JM, Trullen J, Immediato S, Schneider F. Front and backstages of the diminished routinization of innovations: what innovation research makes public and organizational research finds behind the scenes. Strateg Entrep J 2008 Dec;1(3-4):295-314. [doi: 10.1002/sej.21]
40. Pinch TJ, Bijker WE. The social construction of facts and artefacts: or how the sociology of science and the sociology of technology might benefit each other. Soc Stud Sci 1984 Aug 01;14(3):399-441. [doi: 10.1177/030631284014003004]
41. Baron LF, Gomez R. The associations between technologies and societies: the utility of actor-network theory. Sci Technol Soc 2016 Jun 09;21(2):129-148. [doi: 10.1177/0971721816640615]
42. Fuglsang L, Sørensen F. The balance between bricolage and innovation: management dilemmas in sustainable public innovation. Serv Indust J 2010;31(4):581-595.
43. Pols J. Care at a Distance. Amsterdam: Amsterdam University Press; 2012.
44. Krontoft A, Fuglsang L, Kronborg H. Innovation activity among nurses: the translation and preliminary validation of the Bricolage Measure—a mixed-method study. Nord J Nurs Res 2017 Sep 29;38(3):151-159. [doi: 10.1177/2057158517733931]
45. Star S, Strauss A. Layers of silence, arenas of voice: the ecology of visible and invisible work. CSCW 1999;8(1):9-30.
46. Procter R, Greenhalgh T, Wherton J, Sugarhood P, Rouncefield M, Dewsbury G. The ATHENE Project: the importance of bricolage in personalising assisted living technologies. Int J Integr Care 2013;13(7):1.
47. Daatland S, Veenstra M, Kjelvik J, Otnes B, Aksøy H. [Bærekraftig omsorg? Familien, velferdsstaten og aldringen av befolkningen]. Oslo: Norsk institutt for forskning om oppvekst; 2012.
48. Genet N. European Observatory on Health, Policies, Home Care Across Europe: Current Structure and Future Challenges. Observatory Studies Series. Vol. 27. Copenhagen: World Health Organization; 2012.
49. Butterfield LD, Borgen WA, Amundson NE, Maglio AT. Fifty years of the critical incident technique: 1954-2004 and beyond. Qual Res 2016 Aug 17;5(4):475-497. [doi: 10.1177/1468794105056924]
50. Fuglsang L. The critical incident technique and everyday innovation. In: Soensen F, editor. Service Innovation Research Methods. Cheltenham: Edward Elgar; 2017.
51. Chell E, Pittaway L. A study of entrepreneurship in the restaurant and cafe industry: exploratory work using the critical incident technique as a methodology. Int J Hosp Manag 1998 Mar;17(1):23-32. [doi: 10.1016/S0278-4319(98)00006-1]
52. Schluter J, Seaton P, Chaboyer W. Critical incident technique: a user's guide for nurse researchers. J Adv Nurs 2008 Jan;61(1):107-114. [doi: 10.1111/j.1365-2648.2007.04490.x] [Medline: 18173737]
53. Chell E. The critical incident technique: philosophical underpinnings, method and application to a case of small business failure. In: Chell E, editor. Handbook of Research on Small Business and Entrepreneurship. Cheltenham: Edward Elgar; 2014:106-129.
54. Tjora A. [Kvalitative forskningsmetoder: praksis]. Oslo: Gyldendal akademisk; 2012.
55. Strauss A, Corbin J. Grounded theory: methodology. In: Strauss A, editor. Grounded Theory in Practice. Thousand Oaks: Sage; 1997:273-285.
56. Strauss A, Corbin J. Basics of Qualitative Research: Grounded Theory Procedures and Techniques. Newbury Park: Sage; 1990.
57. Patton MQ. Enhancing the quality and credibility of qualitative analysis. Health Serv Res 1999 Dec;34(5 Pt 2):1189-1208 [FREE Full text] [Medline: 10591279]

58. Porter EJ. Wearing and using personal emergency respone system buttons. J Gerontol Nurs 2005 Oct;31(10):26-33. [Medline: 16262088]
59. Porter EJ, Ganong LH, Matsuda S. Intentions of older homebound women with regard to reaching help quickly. West J Nurs Res 2013 Jan;35(1):6-23. [doi: 10.1177/0193945911428482] [Medline: 22146885]
60. Oudshoorn N, Pinch T. How users and non-users matter. In: Oudshoorn N, editor. How Users Matter: The Co-Construction of Users and Technology. Cambridge: MIT Press; 2005.

Abbreviations

CIT: critical incidence technique
STS: science and technology studies

Edited by G Eysenbach; submitted 06.02.18; peer-reviewed by A Essén, B Renn; comments to author 15.03.18; revised version received 28.06.18; accepted 17.07.18; published 19.10.18.

Please cite as:
Stokke R
Older People Negotiating Independence and Safety in Everyday Life Using Technology: Qualitative Study
J Med Internet Res 2018;20(10):e10054
URL: http://www.jmir.org/2018/10/e10054/
doi:10.2196/10054
PMID:30341049

Original Paper

Improving the Understanding of Test Results by Substituting (Not Adding) Goal Ranges: Web-Based Between-Subjects Experiment

Aaron M Scherer[1], PhD; Holly O Witteman[2,3,4], PhD; Jacob Solomon[5], PhD; Nicole L Exe[5], MPH; Angela Fagerlin[6,7], PhD; Brian J Zikmund-Fisher[5,8,9], PhD

[1]Department of Internal Medicine, University of Iowa, Iowa City, IA, United States
[2]Department of Family and Emergency Medicine, Université Laval, Quebec City, QC, Canada
[3]Office of Education and Professional Development, Faculty of Medicine, Université Laval, Quebec City, QC, Canada
[4]Population Health and Optimal Health Practices Research Unit, Research Centre of the CHU de Québec-Université Laval, Quebec City, QC, Canada
[5]Center for Bioethics and Social Sciences in Medicine, University of Michigan, Ann Arbor, MI, United States
[6]Department of Population Health Sciences, University of Utah, Salt Lake City, UT, United States
[7]Salt Lake City VA Informatics Decision-Enhancement and Analytic Sciences (IDEAS) Center for Innovation, Salt Lake City, UT, United States
[8]Department of Health Behavior and Health Education, University of Michigan, Ann Arbor, MI, United States
[9]Department of Internal Medicine, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Aaron M Scherer, PhD
Department of Internal Medicine
University of Iowa
200 Hawkins Drive
Iowa City, IA, 52242
United States
Phone: 1 3193356707
Email: aaron-scherer@uiowa.edu

Abstract

Background: Most displays of laboratory test results include a standard reference range. For some patients (eg, those with chronic conditions), however, getting a result within the standard range may be unachievable, inappropriate, or even harmful.

Objective: The objective of our study was to test the impact of including clinically appropriate goal ranges outside the standard range in the visual displays of laboratory test results.

Methods: Participants (N=6776) from a demographically diverse Web-based panel viewed hypothetical glycated hemoglobin (HbA_{1c}) test results (HbA_{1c} either 6.2% or 8.2%) as part of a type 2 diabetes management scenario. Test result visual displays included either a standard range (4.5%-5.7%) only, a goal range (6.5%-7.5%) added to the standard range, or the goal range only. The results were displayed in 1 of the following 3 display formats: (1) a table; (2) a simple, two-colored number line (simple line); or (3) a number line with diagnostic categories indicated via colored blocks (block line). Primary outcome measures were comprehension of and negative reactions to test results.

Results: While goal range information did not influence the understanding of HbA_{1c}=8.2% results, the goal range only display produced higher levels of comprehension and decreased negative reactions to HbA_{1c}=6.2% test results compared with the no goal range and goal range added conditions. Goal range information was less helpful in the block line condition versus the other formats.

Conclusions: Replacing the standard range with a clinically appropriate goal range could help patients better understand how their test results relate to their personal targets.

(J Med Internet Res 2018;20(10):e11027) doi:10.2196/11027

KEYWORDS

decision making, education of patients, electronic health record, computer graphics, clinical laboratory information systems

Introduction

In an effort to facilitate greater patient involvement in the management of their health, hospitals and health care systems have increasingly provided patients with access to their electronic health records (EHRs) [1]. However, simply providing health information, such as laboratory test results, is often insufficient in enabling patients to understand, much less utilize, this information. Test results are commonly presented in a table format, which leaves a sizable minority of people having difficulty with the seemingly simple task of identifying whether their test result falls within the standard range [2]. Furthermore, even when people can correctly identify the location of their test result in reference to the standard range, they tend to view the risk associated with their test value in a dichotomous fashion—with results within the standard range being viewed as “good” and results outside of the standard range as “bad”—without sensitivity to the fact that risk usually changes in a linear or exponential fashion [2].

Individuals who manage chronic conditions face an additional barrier to understanding and effectively using their test results: inappropriate reference ranges. The standard range commonly presented as part of test result communications represents the distribution of values commonly observed in a healthy population [3-5]. In some chronic disease situations, however, the practical target range that the patient and clinician are trying to reach may be substantially different from the standard range. For example, the standard range for glycated hemoglobin (HbA_{1c}) is generally 4.5%-5.7%, but a common recommendation for patients with type 2 diabetes is to aim to have their HbA_{1c} values below 7%. Furthermore, there is evidence that aggressively managing type 2 diabetes (HbA_{1c} goal: <6.0%) in older individuals results in increased mortality compared with standard therapy (HbA_{1c} goal: 7.0%-7.9%) [6-9], suggesting that at least some patients (eg, those experiencing frequent hypoglycemia) may need to be told that their HbA_{1c} values are lower than advisable. Even in situations where the patient may not be physically harmed by trying to reach the standard range, if the standard range is not realistically achievable, patients may feel justifiably frustrated and discouraged. This could lead to decreased motivation for self-management or the pursuit of alternative therapies in an effort to achieve the unachievable. In an attempt to avoid these potential harms, health care providers frequently discuss goal ranges with their patients that may be more realistic for a person with their condition. Goal ranges may also change with new evidence or changing life circumstances; thus, it may be important to have new ways to communicate these goal ranges via the patient portal.

In addition to the use of clinically appropriate goal ranges, use of visual displays could help increase patient sensitivity to variations among out-of-range results. In a previous study by our research group, we tested the impact of presenting laboratory test results via 3 number line formats versus a standard table format on participants’ sense of urgency and desire to contact their health provider [10]. Compared with participants in the table condition, participants in the 3 number line displays had reduced perceived urgency and desire to contact their health provider for test result values outside of, but near, the standard range. Furthermore, the use of visual displays did not affect participants’ perceived urgency and desire to contact their health provider about more extreme test values.

These issues raise the question of how can test results be communicated to patients in ways that help them better understand how their result compares to the target range most relevant to their self-management and treatment decision making. To the best of our knowledge, there has been no research examining whether and how individual- or disease-specific goal range information should be incorporated into the returned laboratory test results for patients such as these. Inclusion or exclusion of different combinations of these reference standards might improve patients’ comprehension of the test value and reduce unnecessary negative reactions, such as discouragement or urgency to contact their health care provider when urgency is unnecessary.

We conducted a Web-based experiment in which respondents imagined receiving HbA_{1c} test results through an EHR patient portal as part of the ongoing management of their type 2 diabetes. This study was designed to answer four key questions:

- Does the inclusion of goal range information improve comprehension of the test results?
- Does the inclusion of goal range information reduce unnecessary negative reactions to test results that are outside of the standard range, but near their goal range?
- Is it better to include the goal range information in addition to, or in place of, the standard range?
- Does the display format (eg, table vs visual number line) change the impact (if any) of including goal range information in the test result display?

Utilizing the principle “less is more,” which has been shown to apply in health communication [11-14], we hypothesized that the goal information would have the largest improvements in comprehension and reducing unnecessary negative reactions when the goal range was the only reference category (ie, conditions where the standard range and any other risk categories are absent). We also hypothesized that the impact of goal information would be most effective for values nearer to, but still outside of, the standard range because higher test values would be comparatively easier to interpret without additional information.

Methods

Setting

Data were collected through Qualtrics survey software (Qualtrics; Provo, UT) from a nationwide sample of US adults through Survey Sampling International (SSI). Participants were recruited over a 2-month period from August to October 2015.

Sample

Participant eligibility was determined through SSI using a probability-weighted random process based on sample requirements. We established quotas on respondent age (33% aged 21-39 years, 33% aged 40-49 years, and 33% aged ≥60 years), gender (50% females), and race or ethnicity (14%

African American, 14% Hispanic, and 4% Asian American people) to approximate the distribution of these characteristics in the US population. However, we oversampled individuals with diabetes to ensure that we did not have an overly healthy sample and to evaluate whether experience managing diabetes moderated the impact of the goal presentation format. SSI participants were routed to the survey via the sampling algorithm until all quotas were achieved.

Design and Procedure

Participants were asked to imagine that they had recently visited their doctor's office to discuss the management of their type 2 diabetes, during which their doctor had highlighted that people with type 2 diabetes should try to have HbA_{1c} values within a target or goal range of 6.5%-7.5%. Participants were then asked to imagine that in the intervening 3 months, they did their best to follow their doctor's recommendations (eg, exercising regularly and eating healthy). Then, 3 months prior to their next appointment, the patients underwent some blood tests and viewed the results of these tests a day later via a Web-based EHR portal.

We tested 3 between-subjects factors (varied independently) to examine the impact of including goal range information across different presentation formats on patient reactions to their test results. Figure 1 shows examples of different levels of each factor. The first factor was goal presentation. Approximately one-third of participants were randomly presented with a test result display with no goal range (standard range only condition), although the goal range information was described in the scenario text. The remaining participants received visual displays with the goal range included, either in addition to the standard range (goal range added condition) or with the goal range presented instead of the standard range (goal range only condition). The goal range was chosen in consultation with clinicians on our research team who care for patients with type 2 diabetes to represent a realistic and clinically appropriate target range for most individuals with type 2 diabetes.

The second factor was HbA_{1c} test value. Participants were randomly presented with an HbA_{1c} test value of 6.2%, which fell between the standard range and the goal range, or a value of 8.2%, which was higher than both the standard and goal ranges. The third factor was the display format**.** The HbA_{1c} test result was randomly presented via 1 of the 3 formats. The table format presented information via text in a table, the format typically used in EHRs. The simple line format was a gray number line, except for a green range labeled "standard range." The block line format was a number line divided into differently colored diagnostic ranges. The cutoffs for the diagnostic categories were determined in consultation with clinician team members and differed for the "goal range only" condition compared with the "standard range only" and "goal range added" conditions to reflect the differences in hypoglycemia risk for individuals with type 2 diabetes compared with the general population. We reviewed our designs in color vision difference simulators to ensure that the different colors were distinguishable for people with color vision differences. For the "block line" plus "goal range only" combinations, the "Borderline Low" label was represented differently (dropped down, with dotted line connection) as a result of the label being greater in length than the range on the number line.

Measures

Comprehension

We included 2 measures to assess how well participants understood their test result in relation to their goal range. For the relative location measure, we asked, "Where was your test result compared to your goal range?" with "higher than the goal range," "within the goal range," "below the goal range," and "I don't know" as response options. For the future location measure, we asked, "At your next test, what do you think your next test result should be, as compared to this test result?" using a 9-point Likert scale response option with "A lot lower" and "A lot higher" as the anchor labels and "About the same" as the midpoint label. "I don't know" was also included as an additional response option.

Reactions to Test Result

We included 2 measures to assess participant reactions to their test result: one measuring how discouraged they would be by their test result and one assessing whether and when they would contact their doctor about their test result. For the discouraged measure, we asked, "How discouraged or encouraged do you feel about this test result?" using a 6-point Likert scale response option with "Very discouraged" and "Very encouraged" as the anchor labels with an additional "I don't know" response option. For the urgency measure, we asked, "How soon do you need to speak to your doctor regarding these results?" with "Immediately," "Within a few weeks," "At your next appointment in 3 months," and "I don't need to speak to my doctor about these results" as response options.

Demographics

We asked participants about their age, gender, race and ethnicity, education, and whether they have diabetes, and if so, what type.

Data Management

All data were collected anonymously so that the researchers had no way to learn the identity of the participants. A unique identification number provided by SSI was contained in the redirected URL, which identified participants and prevented them from completing the study multiple times. This study was deemed exempt by the University of Michigan Health Sciences and Behavioral Sciences Institutional Review Board.

Figure 1. Goal presentation and display formats for 6.2% glycated hemoglobin (hemoglobin A_{1c}) test value; labels indicate the display format and goal presentation.

Data Analysis

Recoding of Measures

Responses to the relative location measure were recoded as “1” to indicate a correct response if participants responded “below the goal range” in the 6.2% HbA_{1c} test result condition or “higher than the goal range” in the 8.2% HbA_{1c} test result condition. All other responses were recoded as “0” to indicate a failure to know where their test value was in relation to the goal range. To assess whether participants had the gist of where their next test value should be, future location responses were recoded as “1” if they were above the midpoint of the scale in the 6.2% HbA_{1c} test result condition and below the midpoint in the 8.2% condition. All other responses were recoded as “0.” The results are substantially the same, if not stronger (ie, larger effect sizes), if “about the same” is coded as “1”. The one exception is that having diabetes is associated with an increased comprehension of the future location for the goal presentation and display format logistic regression analysis. Responses to the discouraged and urgency measures were reverse coded, such that higher scores indicated greater discouragement and urgency, respectively. We recoded gender (0=male, 1=female), race (0=white, 1=nonwhite), and diabetes status (0=no diabetes, 1=diabetes).

Effects of Goal Presentation

We report percentages for the relative and future location measures and descriptive measures for the discouraged and urgency measures across the different factors. We used chi-square analyses to test for differences in percentages and independent sample *t* tests and one-way analyses of variance (ANOVAs) with post hoc comparisons using Bonferroni corrections for multiple comparisons to compare means. We also report the logistic regression results for the relative and future location measures and ordered logistic regression results for the discouraged and urgency measures, with age, gender, race, education, and diabetes as covariates to test whether including relevant covariates substantially changed the results from the chi-square test, *t* test, and ANOVA. All analyses were performed using Stata 14, and all tests of significance were 2-sided and used alpha=.05.

Results

Sample Description

Of all the participants who initiated the study, 83.09% (6781/8161) completed it. In addition, 14 responses were dropped due to a reported age <18 years old, and 1 response was dropped due to a reported age of 586. Table 1 presents sample demographic characteristics among the remaining 6766 participants.

Impact of Goal Presentation on Interpretation of Tables

In univariate analyses of participants receiving HbA_{1c}=8.2% in table form, neither goal presentation nor display factors significantly affected any of the outcomes (all *P* values>.07, see Figure 2).

Among participants who received tabular displays of HbA_{1c}=6.2% results (which fell between the standard and goal ranges), however, goal presentation format had a significant impact on comprehension. As shown in Figure 2, receiving explicit goal information (in either form) significantly increased the percentage of participants recognizing that their HbA_{1c}=6.2% value was below the goal range (χ^2_2=126.9, *P*<.001) and stating that their next result should be higher (χ^2_2=36.0, *P*<.001). Furthermore, the effect was larger among participants who viewed a table with the goal range only versus when the goal range was added to the standard range (relative location, goal range only: 51.28%, 201/1130, vs goal range added: 44.16%, 155/1130; χ^2_1=3.8, *P*=.05; future location, goal range only: 46.97%, 186/1137, vs goal range added: 37.39%, 132/1137; χ^2_1=7.0, *P*=.01). Similarly, providing goal information (either format) in table displays reduced discouragement ($F_{2,1071}$=19.38, *P*<.001) and urgency ($F_{2,1131}$=3.09, *P*=.046) compared with no goal displays, although there was no significant difference between the goal range added versus goal only conditions.

The logistic regression analyses of participants receiving test results in table format (Table 2) confirmed significant main effects for HbA_{1c} test value and goal presentation for all 4 outcome measures (all *P* values<.001) with the exception of goal presentation for urgency, which became nonsignificant when controlling for the covariates (χ^2_2=5.5, *P*=.06). Consistent with the pattern seen in Figure 2, there were significant interactions between HbA_{1c} test value and goal presentation for relative location (χ^2_2=62.8, *P*<.001), future location (χ^2_2=11.4, *P*=.003), and discouragement (χ^2_2=7.2, *P*=.03), but not for urgency (χ^2_2=1.2, *P*=.54). In addition, individuals with diabetes had not only a lower likelihood of identifying the relative location but also lower discouragement relative to individuals without diabetes. Being older and female were associated with an increased likelihood of identifying the relative location, but decreased urgency. Additionally, being older was associated with increased discouragement while identifying as female was associated with an increased likelihood of correctly identifying where their next test result should be. Identifying as a person of color (nonwhite) was associated with a decreased likelihood of identifying the relative location. Higher education was associated with not only an improved comprehension of the relative location but also increased urgency.

Table 1. Sample characteristics (N=6766).

Characteristic	Value[a]
Age, mean (SD)	49.1 (15.8)
Gender, n (%)	
Male	3299 (48.88)
Female	3435 (50.90)
Transgender or other	15 (0.22)
Ethnicity, n (%)	
Hispanic (any race)	892 (13.26)
Race[b], n (%)	
White	5294 (78.24)
African American	1002 (14.81)
All other	654 (9.67)
Education, n (%)	
<High school	135 (2.00)
High school only	1065 (15.78)
Some college or trade	2458 (36.41)
Bachelor’s degree	2005 (29.70)
>Bachelor’s degree	1087 (16.10)
Diabetes status, n (%)	
No diabetes	3620 (53.79)
Type 1 diabetes	497 (7.38)
Type 2 diabetes	2613 (38.83)
Goal presentation, n (%)	
Standard range only	2253 (33.30)
Goal range added	2219 (32.80)
Goal range only	2294 (33.90)
Glycated hemoglobin test result value, n (%)	
6.2%	3390 (50.10)
8.2%	3376 (49.90)
Display format, n (%)	
Table	2251 (33.27)
Simple line	2224 (32.87)
Block line	2291 (33.86)

[a] Results reported only for those respondents who completed each question or measure.

[b] Respondents could mark more than one race.

Figure 2. Effect of providing goal range information in table format, by goal presentation type and glycated hemoglobin (HbA_{1c}) test result; asterisks indicate statistically significant differences between the 2 bars. Std range: standard range.

Table 2. Logistic regression and ordered logistic regression results showing predictors of outcome measures, table condition only.

Predictors	Relative location		Future location		Discouraged		Urgency	
	OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Goal presentation								
Standard range only	Reference		Reference		Reference		Reference	
Goal range added	4.98 (3.46-7.17)	<.001	1.69 (1.23-2.32)	<.001	0.51 (0.39-0.67)	<.001	0.77 (0.59-1.00)	.049
Goal range only	6.83 (4.77-9.76)	<.001	2.52 (1.86-3.42)	<.001	0.47 (0.35-0.59)	<.001	0.74 (0.57-0.95)	.02
Glycated hemoglobin (HbA_{1c}) test result value								
6.2%	Reference		Reference		Reference		Reference	
8.2%	12.07 (8.30-17.54)	<.001	4.30 (3.13-5.91)	<.001	4.13 (3.12-4.46)	<.001	2.16 (1.63-2.86)	<.001
Goal × HbA_{1c}								
Goal range added × 8.2%	0.25 (0.15-0.40)	<.001	0.57 (0.37-0.87)	.01	1.40 (0.95-2.05)	.09	1.24 (0.84-1.84)	.28
Goal range only × 8.2%	0.15 (0.09-0.24)	<.001	0.49 (0.32-0.75)	<.001	1.67 (1.15-2.44)	.01	1.16 (0.79-1.71)	.46
Demographics								
Diabetes[b]	0.64 (0.52-0.77)	<.001	1.03 (0.87-1.23)	.72	0.52 (0.44-0.61)	<.001	0.98 (0.84-1.15)	.83
Age[c]	1.02 (1.01-1.03)	<.001	1.00 (0.99-1.00)	.54	1.01 (1.01-1.02)	<.001	0.98 (0.97-0.99)	<.001
Gender[d] (female)	1.74 (1.44-2.11)	<.001	1.03 (1.03-1.47)	.02	1.16 (0.99-1.35)	.06	0.84 (0.72-0.99)	.04
Race[e]	0.70 (0.56-0.87)	.002	1.00 (0.81-1.24)	.99	1.01 (0.84-1.22)	.92	1.08 (0.90-1.32)	.39
Education[c]	1.11 (1.05-1.17)	<.001	1.03 (0.98-1.08)	.24	0.99 (0.94-1.03)	.57	1.08 (1.04-1.13)	<.001
Constant	0.03 (0.02-0.05)	<.001	0.28 (0.16-0.47)	<.001	N/A[f]	N/A	N/A	N/A

[a] OR: odds ratio.

[b] Diabetes (0=no, 1=yes).

[c] Age and education treated as continuous variables.

[d] Gender (0=male, 1=female).

[e] Race (0=white, 1=nonwhite).

[f] N/A: not applicable.

Impact of Display Format on Goal Presentation: Glycated Hemoglobin 6.2% Condition Only

Given that providing goal information to participants receiving test results via tables only influenced outcomes among those viewing HbA_{1c}=6.2% results, we focused only on these conditions when comparing optimal formats (ie, table vs simple line vs blocks line) for presenting goal information. As shown in Figure 3, the overall pattern of goal presentation on the understanding and interpretation of HbA_{1c}=6.2% results in the simple line and block line conditions mirrored the pattern discussed above for the table format: providing goal information (in any format) increased the percentage of participants recognizing that their HbA_{1c} value was below the goal range, wanting their next result to be higher and experiencing less discouragement; however, there were minor differences with the block line design. Compared with the table and simple line designs where the goal range was added, participants in the block line condition exhibited less comprehension of their goal location (χ^2_2=13.9, P<.001) and where their next test result should be (χ^2_2=19.4, P<.001) as well as exhibited a greater discouragement $F_{2,1026}$=11.42, P<.001).

The logistic regression analyses (Table 3) revealed the main effects of goal presentation (all *P* values ≤.002), with participants in the goal range added and goal range only conditions having higher comprehension and less discouragement and urgency than participants in the no goal condition and display type (all *P* values≤.002 for relative location, future location, and discouragement measures). More interestingly, there were also significant interactions between goal presentation and display format for relative location (χ^2_4=11.6, P=.02) and future location (χ^2_4=22.2, P<.001), but not for discouragement (χ^2_4=8.0, P=.09) or urgency (χ^2_4=3.2, P=.53). As noted earlier, the block line design seemed to interfere with the efficacy of including goal information when the standard range was also present (Figure 3).

Figure 3. Effects of presenting goal information to patients viewing glycated hemoglobin 6.2% test result: goal presentation and display format; asterisks indicate statistically significant differences between the 2 bars. Std Range: standard range.

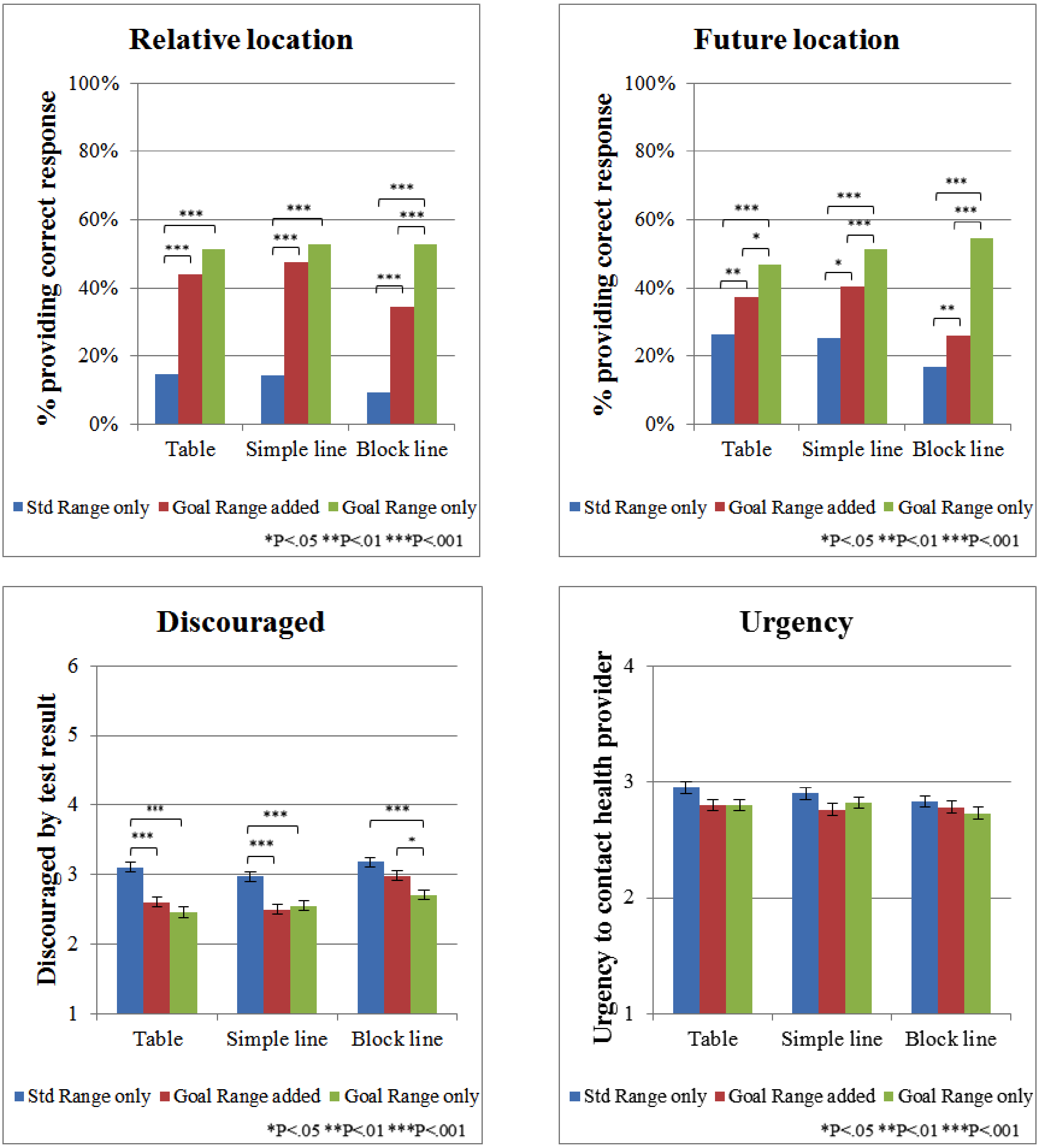


Table 3. Logistic regression results showing goal presentation, presentation format, and demographics as predictors of outcome measures, 6.2% glycated hemoglobin test value condition only.

Predictors	Relative location		Future location		Discouraged		Urgency	
	OR[a] (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Goal presentation								
Standard range only	Reference		Reference		Reference		Reference	
Goal range added	4.85 (3.38-6.96)	<.001	1.73 (1.25-2.39)	<.001	0.45 (0.34-0.59)	<.001	0.74 (0.56-0.96)	.03
Goal range only	6.6 (4.65-9.44)	<.001	2.65 (1.94-3.63)	<.001	0.39 (0.30-0.51)	<.001	0.72 (0.55-0.93)	.01
Presentation format								
Table	Reference		Reference		Reference		Reference	
Simple line	1.02 (0.68-1.54)	.92	1.01 (0.73-1.42)	.93	0.85 (0.65-1.11)	.23	0.92 (0.71-.20)	.55
Block line	0.61 (0.39-0.97)	.04	0.60 (0.41-0.86)	.01	1.17 (0.90-1.53)	.25	0.83 (0.64-1.08)	.16
Goal × presentation								
Goal range added × simple line	1.17 (0.70-1.95)	.54	1.17 (0.73-1.85)	.51	1.16 (0.78-1.70)	.46	0.99 (0.68-1.46)	.97
Goal range added × block line	1.07 (0.62-1.86)	.82	1.02 (0.63-1.67)	.93	1.56 (1.07-2.28)	.02	1.22 (0.83-1.78)	.31
Goal range only × simple line	0.99 (0.60-1.64)	.97	1.19 (0.76-1.86)	.45	1.36 (0.94-1.97)	.11	1.20 (0.83-1.74)	.34
Goal range only × block line	1.74 (1.01-2.99)	.045	2.32 (1.45-3.71)	<.001	1.29 (0.89-1.86)	.18	1.09 (0.76-1.58)	.64
Demographics								
Diabetes[b]	0.64 (0.55-0.75)	<.001	0.77 (0.66-0.89)	<.001	0.31 (0.27-0.35)	<.001	0.74 (0.65-0.84)	<.001
Age[c]	1.01 (1.01-1.02)	<.001	0.97 (0.96-0.97)	<.001	1.00 (0.99-1.00)	.049	0.97 (0.97-0.98)	<.001
Female gender[d]	1.56 (1.33-1.82)	<.001	0.90 (0.77-1.05)	.19	1.04 (0.91-1.18)	.57	0.79 (0.69-0.89)	<.001
Race[e]	0.76 (0.63-0.92)	.01	1.02 (0.85-1.22)	.85	0.96 (0.82-1.12)	.62	1.20 (1.03-1.40)	.02
Education[c]	1.11 (1.06-1.16)	<.001	1.04 (1.00-1.09)	.07	0.98 (0.95-1.02)	.29	1.03 (0.99-1.06)	.14
Constant	0.04 (0.03-0.07)	<.001	1.38 (0.87-2.19)	.17	N/A[f]	N/A	N/A	N/A

[a] OR: odds ratio.

[b] Diabetes (0=no, 1=yes).

[c] Age and education treated as continuous variables.

[d] Gender (0=male, 1=female)

[e] Race (0=white, 1=nonwhite)

[f] N/A: not applicable.

Demographic covariates remained significant predictors across the 4 outcome measures (see Table 3). People with diabetes were less likely to identify the relative location of their result, but they also had less discouragement and urgency compared with individuals without diabetes. Age produced inconsistent effects, with a high comprehension of relative location but low comprehension of future location along with lower discouragement and urgency. Identifying as female or white were both associated with an increased comprehension of relative location and with decreased urgency. Education was associated with increased comprehension of relative location.

Interaction Analysis of the Impact of Diabetes Status: Glycated Hemoglobin 6.2% Condition Only

The regression results presented in Table 3 showed consistent main effects on comparing participants who have diabetes in real life versus those who did not. To explore whether diabetes status might interact with optimal display formats, we performed additional regression analyses including interaction terms based on diabetes status. These additional logistic regression and ordered logistic regression results revealed a significant interaction between diabetes status and goal presentation for comprehension of the relative and future locations (all *P* values<.001), but not for discouragement and urgency (all *P* values>.15; Multimedia Appendix 1). For comprehension of relative location, the overall relationship between the effect of goal presentation did not change based on whether someone had diabetes (no goal: 80/552, 14.49% vs goal range added: 171/476, 35.92% and goal range only: 219/504, 43.45%; *P*<.001) or not (no goal: 65/578, 11.25% vs goal range added: 278/595, 46.72% and goal range only: 385/646, 59.60%; *P*<.001); the effects were just more exaggerated for people without diabetes. For comprehension of the future location, there were significant differences between all 3 goal presentation conditions for participants without diabetes (no goal: 132/579, 22.80% vs goal range added: 243/604, 40.23% and goal range only: 364/649, 56.09%; *P*<.001). However, for people with diabetes, comprehension was significantly higher in the goal only condition (225/506, 44.47%) than in no goal (129/555, 23.24%; *P*<.001) and goal range added (127/477, 26.62%; *P*<.001) conditions, while there were no significant differences between the no goal and goal range added condition (*P*=.70).

Discussion

Our data suggest that providing people with test results displays (tabular or visual) that include goal range information can alter their perceptions of their test results in important ways. While perceptions were generally unaffected by format when the result was above both the standard and goal ranges, perceptions were sensitive to format when the result was above the standard range but below the goal range. Comprehension of the below-target nature of this result was higher when goal information was explicitly included in participants' test result tables or visual displays. Furthermore, inclusion of goal information in the display reduced perceived discouragement about the presented results.

Our data also show that removing the standard range and substituting it with a single goal reference range seems superior to simply adding goal range information along with the standard range values. Comprehension was highest and discouragement and urgency were lowest when the goal range information was presented in lieu of the standard range information. This suggests that it is difficult for people to put aside information about the standard range—which is normed based on the total, mostly healthy, population—even when more personalized goal information is easily available. As a result, the inclusion of these standard reference points (which are less relevant in this particular situation) may undermine patients' ability to manage their chronic conditions and may expose them to harm when aggressively trying to achieve test results within the standard range [6-9].

Fundamental principles of both visual design and information evaluability suggest that the dominance of the goal only substitution condition is due to the fact that the inclusion of more than 1 reference range produces confusion about which comparator is most relevant to understanding where the patient's test value should be [15,16]. This argument is bolstered by the fact that among the conditions where goal information was presented in addition to the standard range, comprehension was lowest and discouragement was highest when participants received block design visuals. This design already includes multiple color-coded sections and categorical labels indicating levels of risk, and adding yet another reference range for patients to interpret at the same time was clearly too much for many to handle.

One limitation of our study is the use of a hypothetical scenario. While participants did not receive actual test results, approximately half of our sample had the medical condition described in the scenario (diabetes) and would likely have experience receiving HbA_{1c} test results. While we found the same pattern of results for participants with and without diabetes, participants with diabetes who received HbA_{1c}=6.2% results were less likely to report that their values were too low, but these participants also exhibited decreased discouragement and urgency. One possible explanation for this finding is that their experience with repeatedly being told that their HbA_{1c} goal should be below 7.0% has led them to adopt the standard range as the norm that they should be striving to attain, even when an alternative goal range has been provided. Another possibility is that participants with diabetes were relying on their real-life goal ranges, which may have been different from the one provided in the scenario, or that they recognize that not all persons with type 2 diabetes will experience adverse outcomes with an HbA_{1c} of 6.2%. This explanation may account for the overall smaller percentage of participants with diabetes who were discouraged about their test result or felt a need to contact their health care provider immediately.

As more and more patients receive their test results via Web-based patient portals, it is becoming increasingly important that patients should be able to find their results meaningful and that we do not cause unnecessary distress or discouragement to patients. Current approaches to presenting laboratory test results to patients appear to be particularly problematic for many patients, such as those with chronic conditions, who may have personal target goals that differ from those relevant to healthy adults. For these patients, the standard range commonly shown is not necessarily where we want patient results to be. Providing goal range information in place of the standard range may be one step toward reducing these problems with EHR systems; however, challenging discussions would need to occur regarding the pros and cons of who should determine the goal range information (ie, health systems, EHR or portal vendors, expert panels, individual physicians, and/or patients) or what the goal ranges should represent (eg, broader goals for people with a chronic condition vs individualized goals). More research is needed to determine additional features that may further improve the interpretability of laboratory test results.

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 by a grant from the US Agency for Healthcare Research and Quality to BJZF (R01 HS021681). The funding agreement assured the authors' independence in designing the study; in the collection, analysis, and reporting of the data; and in the decision to submit the article for publication. The authors acknowledge the assistance of Sandeep Vijan, MD, Kenneth Langa, MD, and Beth Tarini, 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 Grace

Bienek in creating the images used in this study. 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.

Multimedia Appendix 1

Logistic regression and ordered logistic regression results showing goal presentation, diabetes status, and demographics as predictors of outcome measures, 6.2% A_{1c} test value condition only.

[PDF File (Adobe PDF File), 194KB - jmir_v20i10e11027_app1.pdf]

References

1. American Hospital Association. Individuals' Ability to Electronically Access Their Hospital Medical Records, Perform Key Tasks is Growing. 2016 Oct 29. URL: http://www.aha.org/research/reports/tw/16jul-tw-healthIT.pdf [accessed 2018-09-14] [WebCite Cache ID 72QqV2lU1]
2. 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]
3. Doumas BT. The evolution and limitations of accuracy and precision standards. Clin Chim Acta 1997 Apr 25;260(2):145-162. [Medline: 9177910]
4. Ozarda Y. Reference intervals: current status, recent developments and future considerations. Biochem Med (Zagreb) 2016;26(1):5-16 [FREE Full text] [doi: 10.11613/BM.2016.001] [Medline: 26981015]
5. Siest G, Henny J, Gräsbeck R, Wilding P, Petitclerc C, Queraltó JM, et al. The theory of reference values: an unfinished symphony. Clin Chem Lab Med 2013 Jan;51(1):47-64. [doi: 10.1515/cclm-2012-0682] [Medline: 23183761]
6. Action to Control Cardiovascular Risk in Diabetes Study Group, Gerstein HC, Miller ME, Byington RP, Goff DC, Bigger JT, et al. Effects of intensive glucose lowering in type 2 diabetes. N Engl J Med 2008 Jun 12;358(24):2545-2559 [FREE Full text] [doi: 10.1056/NEJMoa0802743] [Medline: 18539917]
7. Lipska KJ, Ross JS, Wang Y, Inzucchi SE, Minges K, Karter AJ, et al. National trends in US hospital admissions for hyperglycemia and hypoglycemia among Medicare beneficiaries, 1999 to 2011. JAMA Intern Med 2014 Jul;174(7):1116-1124 [FREE Full text] [doi: 10.1001/jamainternmed.2014.1824] [Medline: 24838229]
8. Tseng C, Soroka O, Maney M, Aron DC, Pogach LM. Assessing potential glycemic overtreatment in persons at hypoglycemic risk. JAMA Intern Med 2014 Feb 01;174(2):259-268. [doi: 10.1001/jamainternmed.2013.12963] [Medline: 24322626]
9. Pogach L, Tseng C, Soroka O, Maney M, Aron D. A Proposal for an Out-of-Range Glycemic Population Health Safety Measure for Older Adults With Diabetes. Diabetes Care 2017 Dec;40(4):518-525 [FREE Full text] [doi: 10.2337/dc16-0953] [Medline: 28325799]
10. 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]
11. Peters E, Dieckmann N, Dixon A, Hibbard JH, Mertz CK. Less is more in presenting quality information to consumers. Med Care Res Rev 2007 Apr;64(2):169-190. [doi: 10.1177/10775587070640020301] [Medline: 17406019]
12. Zikmund-Fisher BJ, Fagerlin A, Ubel PA. A demonstration of "less can be more" in risk graphics. Med Decis Making 2010;30(6):661-671 [FREE Full text] [doi: 10.1177/0272989X10364244] [Medline: 20375419]
13. Zikmund-Fisher BJ, Angott AM, Ubel PA. The benefits of discussing adjuvant therapies one at a time instead of all at once. Breast Cancer Res Treat 2011 Aug;129(1):79-87 [FREE Full text] [doi: 10.1007/s10549-010-1193-4] [Medline: 20945090]
14. Zikmund-Fisher BJ, Fagerlin A, Ubel PA. Improving understanding of adjuvant therapy options by using simpler risk graphics. Cancer 2008 Dec 15;113(12):3382-3390 [FREE Full text] [doi: 10.1002/cncr.23959] [Medline: 19012353]
15. Zikmund-Fisher BJ, Fagerlin A, Ubel PA. "Is 28% good or bad?" Evaluability and preference reversals in health care decisions. Med Decis Making 2004;24(2):142-148. [doi: 10.1177/0272989X04263154] [Medline: 15090100]
16. Tufte ER. The Visual Display of Quantitative Information. Cheshire, CT: Graphics Press; 2001.

Abbreviations

ANOVA: analysis of variance
EHR: electronic health record
HbA_{1c}: glycated hemoglobin
SSI: Survey Sampling International

Edited by G Eysenbach; submitted 15.05.18; peer-reviewed by C Schäfer, HC Kum; comments to author 27.06.18; revised version received 05.07.18; accepted 10.07.18; published 19.10.18.

Please cite as:
Scherer AM, Witteman HO, Solomon J, Exe NL, Fagerlin A, Zikmund-Fisher BJ
Improving the Understanding of Test Results by Substituting (Not Adding) Goal Ranges: Web-Based Between-Subjects Experiment
J Med Internet Res 2018;20(10):e11027
URL: http://www.jmir.org/2018/10/e11027/
doi:10.2196/11027
PMID:30341053

Original Paper

Television Advertising and Health Insurance Marketplace Consumer Engagement in Kentucky: A Natural Experiment

Paul R Shafer[1], MA; Erika Franklin Fowler[2,3], PhD; Laura Baum[3], MURP; Sarah E Gollust[4], PhD

[1]Department of Health Policy and Management, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

[2]Government Department, Wesleyan University, Middletown, CT, United States

[3]Wesleyan Media Project, Wesleyan University, Middletown, CT, United States

[4]Division of Health Policy and Management, School of Public Health, University of Minnesota, Minneapolis, MN, United States

Corresponding Author:
Paul R Shafer, MA
Department of Health Policy and Management
Gillings School of Global Public Health
University of North Carolina at Chapel Hill
135 Dauer Drive
Campus Box 7411
Chapel Hill, NC, 27599
United States
Phone: 1 919 260 2711
Email: pshafer@unc.edu

Abstract

Background: Reductions in health insurance enrollment outreach could have negative effects on the individual health insurance market. Specifically, consumers may not be informed about the availability of coverage, and if some healthier consumers fail to enroll, there could be a worse risk pool for insurers. Kentucky created its own Marketplace, known as kynect, and adopted Medicaid expansion under the Affordable Care Act, which yielded the largest decline in adult uninsured rate in the United States from 2013 to 2016. The state sponsored an award-winning media campaign, yet after the election of a new governor in 2015, it declined to renew the television advertising contract for kynect and canceled all pending television ads with over a month remaining in the 2016 open enrollment period.

Objective: The objective of this study is to examine the stark variation in television advertising across multiple open enrollment periods in Kentucky and use this variation to estimate the dose-response effect of state-sponsored television advertising on consumer engagement with the Marketplace. In addition, we assess to what extent private insurers can potentially help fill the void when governments reduce or eliminate television advertising.

Methods: We obtained television advertising (Kantar Media/Campaign Media Analysis Group) and Marketplace data (Kentucky Health Benefit Exchange) for the period of October 1, 2013, through January 31, 2016, for Kentucky. Advertising data at the spot level were collapsed to state-week counts by sponsor type. Similarly, a state-week series of Marketplace engagement and enrollment measures were derived from state reports to Centers for Medicare and Medicaid Services. We used linear regression models to estimate associations between health insurance television advertising volume and measures of information-seeking (calls to call center; page views, visits, and unique visitors to the website) and enrollment (Web-based and total applications, Marketplace enrollment).

Results: We found significant dose-response effects of weekly state-sponsored television advertising volume during open enrollment on information-seeking behavior (marginal effects of an additional ad airing per week for website page views: 7973, visits: 390, and unique visitors: 388) and enrollment activity (applications, Web-based: 61 and total: 56).

Conclusions: State-sponsored television advertising was associated with nearly 40% of unique visitors and Web-based applications. Insurance company television advertising was not a significant driver of engagement, an important consideration if cuts to government-sponsored advertising persist.

(J Med Internet Res 2018;20(10):e10872) doi:10.2196/10872

KEYWORDS

advertising; Affordable Care Act; enrollment

Introduction

In late August 2017, the Trump administration announced plans to cut federal funding for Marketplace advertising by 90% for the 2018 open enrollment period, including the elimination of all television advertising [1]. This was the latest upheaval for the Affordable Care Act (ACA) in a tumultuous first year of the Trump presidency, which included an executive order to "minimiz[e] the economic burden" of the law, unplanned advertising reductions at the end of the 2017 open enrollment period, failed legislative efforts to repeal and replace, shortening of the open enrollment period, cuts to enrollment assistance, discontinuation of cost-sharing reduction payments, and a repeal of the individual mandate included in the tax bill signed into law in December [2-6]. The Marketplace, as the combination of state-based and federally run health insurance exchanges created by the ACA is known, provides consumers with the opportunity to compare plans, determine eligibility for financial assistance and Medicaid, and enroll in coverage. The ACA relied on the new Marketplace, the expansion of the Medicaid program (among states that chose to do so) to low-income childless adults, and extending insurance coverage eligibility for youth up to age 26 on their parents' plans as the main mechanisms to expand access to health insurance in the United States to approximately 20 million people [7].

Thus far, limited evidence on the impact of Marketplace advertising on enrollment and related outcomes suggests that advertising works, though this evidence is not causal. Higher exposure to advertising has been associated with improved perceptions of and knowledge about the ACA. Among a sample of low-income adults in Arkansas, Kentucky, and Texas, those reporting greater exposure to positive advertising about the law were significantly more likely to say that the "ACA helped me" [8]. In a nationally representative sample, those with higher volumes of insurance-related advertising in their media market were significantly more likely to report feeling informed about the ACA and have more positive views of the law [9]. The two published studies to date that have examined the relationship between the dose of Marketplace advertising and enrollment-related outcomes have found that counties with more advertising saw larger decreases in their uninsured rate and a greater likelihood of shopping for and enrolling in a Marketplace plan during the 2014 open enrollment period [10,11]. California has credited aggressive marketing with higher take-up of coverage and lower risk scores, providing a substantial return on investment in terms of lower premiums [12]. As of yet, however, no published research has examined how changes in advertising volume from multiple sponsors over time and within open enrollment periods correspond with consumer engagement with the Marketplace.

Kentucky is an important case for examining these marketing and consumer engagement dynamics. Kentucky created its own Marketplace, known as kynect, and adopted Medicaid expansion under the ACA, which yielded the largest decline in adult uninsured rate in the country from 2013 to 2016 [13]. The state sponsored an award-winning media campaign and also received national recognition for its implementation of the information technology infrastructure for kynect [14,15]. Yet, after the election of a new governor in 2015, state politics led to changes in kynect messaging and promotion that foreshadowed what was to come at the federal level. Specifically, the Bevin administration declined to renew the advertising contract for kynect and canceled all pending television ads with over a month remaining in the 2016 open enrollment period [16]. The objective of this study is to examine the stark variation in advertising across multiple open enrollment periods in Kentucky and use this variation to estimate the dose-response effect of state-sponsored advertising on consumer engagement with the Marketplace. In addition, we assess to what extent private insurers can potentially help fill the void when governments reduce or eliminate television advertising.

Methods

Data

We used data from Kantar Media/Campaign Media Analysis Group (CMAG) to measure the volume of local broadcast and national cable television advertising for health insurance and from the Office of the Kentucky Health Benefit Exchange (KHBE) to describe consumer engagement with kynect. The Kantar Media/CMAG data—obtained through collaboration with the Wesleyan Media Project—track airings of individual televised advertisements, including date, time, sponsor, station, and media market, for the period of October 1, 2013, through January 31, 2016, spanning the 2014 to 2016 open enrollment periods. We obtained these data for all 10 media markets in Kentucky, including border markets that reach across state lines (only 2 markets are fully contained within the state). We categorized advertisements into one of 6 sponsor types: (1) kynect; (2) healthcare.gov; (3) insurance companies (eg, Aetna and Cigna); (4) insurance agencies (eg, Healthmarkets Insurance Agency); (5) nonprofits; and (6) other state governments (to capture ads from neighboring states aired in Kentucky). Multimedia Appendix 1 provides a list of ad sponsors and their assigned sponsor type. We identified healthcare.gov advertising through health insurance ads paid for by the US Department of Health and Human Services. A small percentage (565/10,089, 5.6%) of these ad airings were not specific to healthcare.gov, but nearly all either mentioned the "health care law" or explicitly tagged healthcare.gov. Only 0.5% (46/10,089) of these airings had no mention of the ACA at all (ads for Medicare open enrollment). We collapsed the data by sponsor type and week (Sunday to Saturday) in each media market and then calculated a population-weighted average across media markets using media market population estimates from Polidata to provide a state-level estimate of ads aired per week for each sponsor type [17].

The KHBE analytic data were derived from a set of reports from the state to the Center for Consumer Information and Insurance Oversight at Centers for Medicare and Medicaid Services that were obtained through state public records request. Of the 122 weeks between October 1, 2013, and January 31, 2016, there was only a single week (February 9-15, 2014) for which a report was not available and 2 other reporting periods (December 13-19, 2015, and June 30-July 28, 2015) for which one of the outcomes that we used in our analysis (calls) was not reported. We used linear interpolation to impute those missing values. The time periods for each report varied and were often longer outside of open enrollment (eg, every 4 weeks compared with every week within the open enrollment period). To maintain consistent measures, we converted data from the longer reporting periods back to a weekly frequency by averaging the incremental activity over the appropriate number of weeks. For example, if there were an incremental gain of 1000 unique visitors over a 4-week reporting period, we converted this back to 4 individual weeks with 250 unique visitors each week. There were a few reporting periods that had a length other than 7 days (eg, 5, 6, and 8) that we could not standardize (eg, the first week in the data, October 1-5, 2013). We addressed this by including the number of days in the reporting period as a covariate in all regression models.

Outcomes

The process of shopping for and enrolling in coverage generally involves visiting the Marketplace website to view plans and pricing with the ability to learn about subsidy eligibility (based on family structure and household income) and then select a plan in which to enroll. For this reason, we focused on both parts of the process. First, we assessed how well television advertising does at driving consumers to visit the kynect website. We could not explicitly examine conversion of those website visits into enrollment, but we assessed how television advertising is associated with application submissions and plan enrollment. We used 7 measures as outcomes to represent 2 domains of consumer engagement with kynect—information-seeking behavior and enrollment activity. Our measures of information-seeking behavior were (1) calls to the kynect call center; (2) page views (number of individual pages viewed); (3) visits (including repeats from the same internet protocol address); and (4) unique visitors (excluding repeats) for the kynect website. Our measures of enrollment activity were (5) number of applications completed through Web; (6) total number of applications completed; and (7) number of individuals enrolled in qualified health plans (net of plan terminations). All outcomes were defined as incremental state-week totals (as described in Data above). The advertising data (reported at the weekly level) were then merged with these outcome data at matching time periods.

Statistical Analysis

We used linear regression models to identify variation in each outcome that was associated with variation in the television advertising volume, with the unit of analysis as a state-week. This approach has been used to analyze consumer engagement with health education campaigns, including the *Tips From Former Smokers* campaign by the Centers for Disease Control and Prevention [18,19]. Each model included the number of ads aired for each of the 6 sponsor types, an indicator for open enrollment periods, and their interaction to allow us to identify the differential effect of advertising during open enrollment. We used a single indicator for all open enrollment periods rather than period-specific indicators (ie, indicators for the 2014, 2015, and 2016 open enrollment periods separately) owing to our limited sample size (122 state-weeks). As such, there may be differential effects related to specific open enrollment periods that we are averaging over. In addition, we included indicators for the weeks of Thanksgiving and Christmas and indicators for the 2 weeks preceding open enrollment (to capture any early plan shopping behavior), the first 2 weeks of open enrollment (to capture early enrollees), and the last 2 weeks of open enrollment (to capture later enrollees). Moreover, we controlled for the number of days in the reporting period as noted above. Average marginal effects were calculated to describe the incremental effect of an additional advertisement per week for each sponsor type. Next, we estimated the share of consumer engagement that was attributable to advertising. To do so, we used linear predictions from each model with the observed advertising during the first 2 open enrollment periods compared with the counterfactual—the absence of kynect or insurance company advertising. This involves using the coefficient estimates from the regression models fitted to the observed data to predict what the outcomes would have been if we set either kynect or insurance company advertising to zero during open enrollment and then comparing those results to the predictions from the model using the actual observed data. All models were estimated with robust standard errors, and *P*<.05 was used as the threshold for statistical significance. Multimedia Appendix 2 shows a correlation matrix for our independent variables. All analyses were conducted in Stata 14.2 [20].

Results

Weekly advertising volume varied widely between and within sponsor types during the study period (October 1, 2013, through January 31, 2016; Figure 1). Insurance companies aired the majority of all advertisements over the 3 open enrollment periods with an average of 169 ad airings per week compared with 61 and 46 airings per week for healthcare.gov and kynect, respectively. Insurance companies aired considerably more ads per week during the 2015 (231 airings) and 2016 (180 airings) open enrollment periods than during 2014 (132 airings). Kynect advertising fell slightly from 2014 (59 airings) to 2015 (52 airings) before a substantial decline for the 2016 open enrollment period (13 airings). In the 2016 open enrollment period, kynect advertising fell from an average of 19 airings per week during the first 9 weeks to none during the final 4 weeks. Furthermore, healthcare.gov advertising fell from a high of 92 airings during the 2014 open enrollment period to only 37 and 23 airings per week during the 2015 and 2016 open enrollment periods, respectively, although these were primarily aired in border markets (Nashville, Cincinnati, Knoxville, and Harrisburg) and likely were not targeting Kentucky specifically. Insurance agencies (20 airings), other state governments (19 airings), and nonprofits (6 airings) each averaged 20 or fewer airings per week over the 3 open enrollment periods.

Figure 1. Weekly advertising volume by sponsor type, Kentucky, October 1, 2013-January 31, 2016.

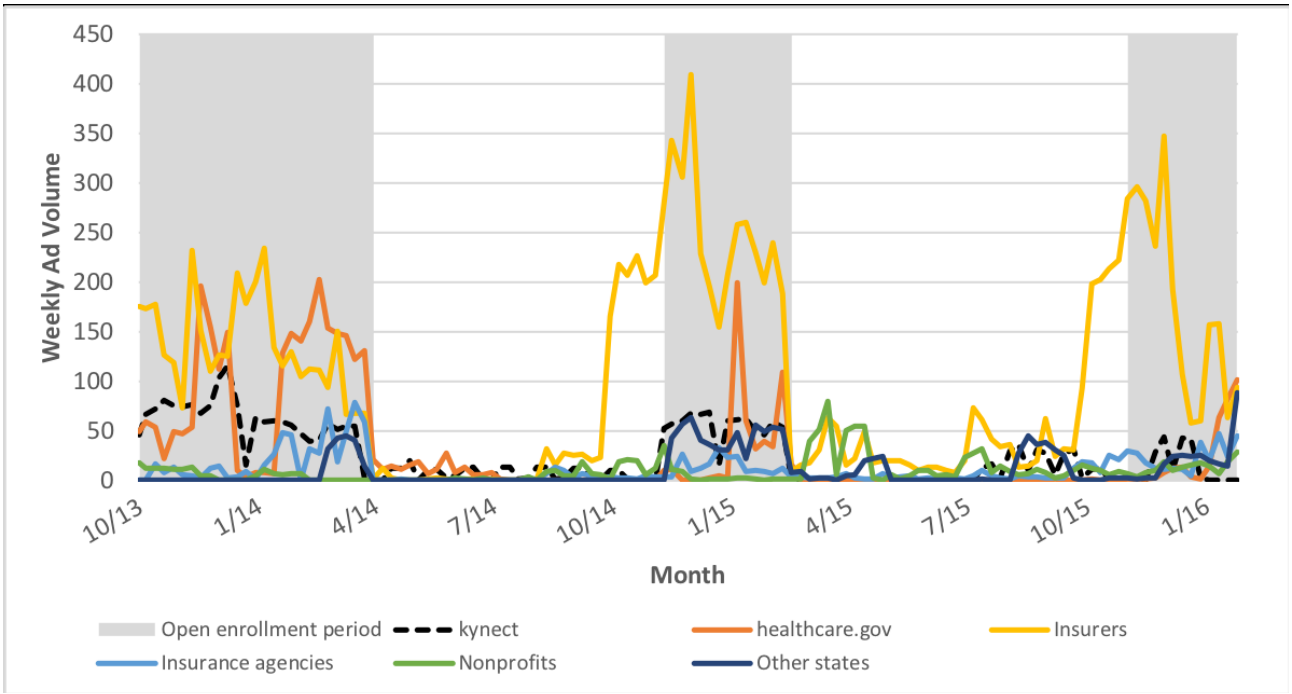


Table 1 presents weekly averages for each information-seeking and enrollment outcome during and outside of open enrollment. Figure 2 shows the weekly trends for calls, unique visitors to the kynect website, total applications, and enrollment. Qualified life events that trigger a special enrollment period (marriage, birth of a child, change in job, change in income, etc) outside of open enrollment provide a baseline level of activity. Calls to the kynect call center were approximately 60% higher during open enrollment, and Web traffic outcomes more than doubled. Similarly, application activity more than doubled through Web and slightly less than doubled in total during open enrollment. Weekly net enrollment in qualified health plans was generally negative outside of open enrollment periods, indicating that disenrollment due to changes in job-related or spousal coverage and nonpayment of premiums outweighed new enrollments outside of open enrollment.

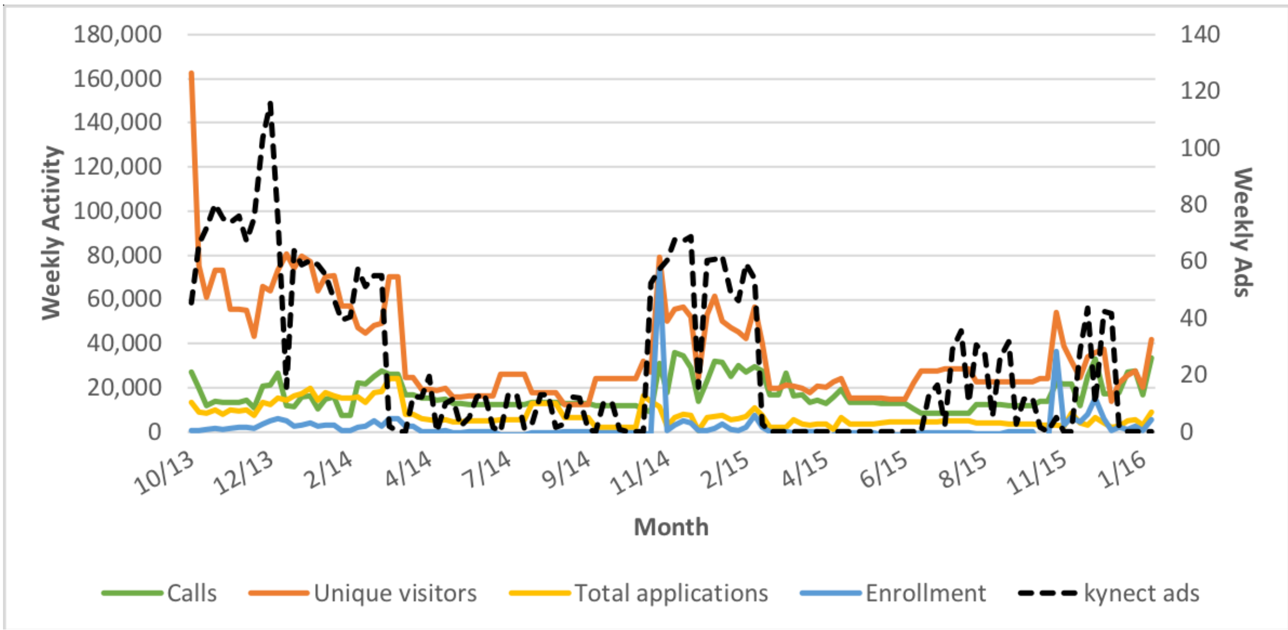
Multimedia Appendix 3 contains linear regression results for our 4 information-seeking behavior outcomes—calls to the kynect call center and page views, visits, and unique visitors for the kynect website. The key measure of dose-response is the marginal effect of advertising volume during open enrollment, which is the sum of the coefficients for the relevant main effect (sponsor type weekly ad volume) and the differential effect (sponsor type weekly ad volume×open enrollment indicator) for each outcome. The marginal effects of kynect advertising during open enrollment are positive and significant for page views, visits, and unique visitors. Specifically, each additional kynect ad airing per week during open enrollment was associated with approximately 8000 additional page views (7972.9, *P*=.001), 400 additional visits (390.2, *P*=.003), and 400 additional unique visitors (387.5, *P*<.001) to the kynect website. There is no evidence of a dose-response effect for advertising by healthcare.gov, insurance companies, and insurance agencies for these outcomes (all marginal effects are not statistically significant for these sponsor types). In contrast, advertising by nonprofits seemed to drive traffic away from kynect resources (calls: −197.6, *P*=.02; page views: −30,061.5, *P*=.03). Advertising by other state governments was associated with increased calls (215.9, *P*<.001) to kynect but no change in Web traffic. Also, we found that the week of Thanksgiving was strongly associated with lower information-seeking activity across all outcomes (calls: −9058.8, *P*=.002; page views: −774,472.7, *P*<.001; visits: −39,814.8, *P*<.001; unique visitors: −23,955.8, *P*<.001) and that the week of Christmas was associated with lower call volume (−8375.9, *P*=.003).

The dose-response effect of kynect advertising on these outcomes was robust to dropping the first week of open enrollment for 2014 (not shown), when the number of unique visitors was at its highest point in the data. The relationship between the kynect advertising volume during open enrollment and calls remained not significant. The relationship (marginal effect) between the kynect advertising volume during open enrollment and page views (7374.7, *P*=.001), visits (347.9, *P*=.005), and unique visitors (357.7, *P*<.001) remained stable and highly significant. We also assessed whether dropping the 2014 open enrollment period altogether (not shown) substantially changed our findings over concerns about the 2014 period being different and driving our overall findings with the Marketplace being new (and, thus, advertising offering novel information). The relationship between kynect advertising volume during open enrollment and calls again was not statistically significant. The relationships (marginal effects) between kynect advertising volume during open enrollment and page views (8705.4, *P*=.02), visits (378.1, *P*<.001), and unique visitors (259.8, *P*<.001) were again similar.

Table 1. Weekly averages for information-seeking and enrollment outcomes by open enrollment, Kentucky, October 1, 2013-January 31, 2016.

Outcome	During open enrollment, mean (range)	Outside of open enrollment, mean (range)
Information-seeking behavior		
Calls	21,348 (7724 to 35,905)	13,045 (8470 to 26,634)
Page views	1,791,512 (471,996 to 3,140,745)	715,252 (473,532 to 1,244,794)
Visits	94,990 (28,312 to 234,711)	42,477 (27,158 to 62,759)
Unique visitors	54,525 (13,881 to 162,774)	21,147 (12,621 to 31,888)
Enrollment activity		
Web-based applications	8032 (400 to 23,378)	3290 (148 to 12,707)
Total applications	10,095 (775 to 24,075)	5179 (949 to 12,792)
Enrollment	5160 (562 to 73,729)	−430 (−6489 to 2600)

Figure 2. Weekly information-seeking and enrollment outcomes, Kentucky, October 1, 2013-January 31, 2016.



Multimedia Appendix 4 contains linear regression results for our 3 enrollment activity outcomes—number of applications completed through Web, total number of applications completed, and net enrollment in qualified health plans. Again, calculating the marginal effects, we found that each additional kynect ad airing per week during open enrollment was associated with approximately 60 additional applications for coverage completed through Web (61.4, *P*=.03) and more than 55 additional total applications completed (55.9, *P*=.03). Each additional ad per week by insurance agencies during open enrollment was associated with increased Web-based applications (111.4, *P*=.02). Each additional ad per week by other state governments during open enrollment was associated with fewer Web-based applications (−122.8, *P*=.001) and in total (−82.1, *P*=.01). All other marginal effects were not statistically significant.

The week of Thanksgiving was associated with a lower volume of applications completed—both Web based (−3434.1, *P*=.03) and in total (−4523.4, *P*<.001)—mirroring what we observed for information-seeking behavior. The percentage of variation in these outcomes explained by our models was considerably lower—ranging from 56% to 58%—than for the information-seeking outcomes (75% to 81%). This is not surprising given that we would not expect advertising to immediately result in enrollment in the same way that we would for information-seeking behavior. Submitting an application for enrollment requires the consumer to provide tax or income information and selecting a Marketplace plan, if initially deemed not to be Medicaid eligible; this decision may require consideration of various plan options over a longer time period and consultation with a spouse.

We again assessed whether the dose-response effect of kynect advertising on these outcomes was robust to dropping the first week of open enrollment for 2014 or the 2014 open enrollment altogether (not shown). For the former, the relationship (marginal effect) of kynect advertising volume during open enrollment with applications completed through Web (57.3, *P*=.04) and total applications (50.8, *P*=.04) was slightly lower in magnitude but similar. We again found no relationship between kynect advertising volume and net enrollment in qualified health plans.

Table 2. Weekly activity attributable to kynect and insurance company advertising, Kentucky, 2014 and 2015 open enrollment periods.

Outcome	Actual, n	Model-based predictions				
		Prediction (with observed advertising), n[a]	Counterfactual (no kynect advertising), n[a]	Change attributable to kynect advertising, n (%)[b]	Counterfactual (no insurance company advertising), n[a]	Change attributable to insurance company advertising, n (%)[c]
Information-seeking behavior						
Page views	1,978,749	1,925,863	1,474,730	451,133 (23)	1,804,879	120,984 (6)
Visits	105,618	101,616	79,538	22,078 (22)	98,061	3555 (3)
Unique visitors	61,838	58,964	37,037	21,927 (37)	55,242	3722 (6)
Enrollment activity						
Web-based applications	9643	9050	5576	3474 (38)	11,759	–2709 (–30)
Total applications	11,832	11,232	8072	3160 (28)	12,516	−1283 (−11)

[a] Represent model-based predictions (using the model results shown in Multimedia Appendix 3 and Multimedia Appendix 4) of the weekly outcomes. These comparisons use the 2014 and 2015 open enrollment periods to estimate how much activity was attributable to advertising. Calls and Marketplace enrollment are not included because neither kynect nor insurance company advertising during open enrollment had a significant effect on these outcomes.

[b] Difference (%) between "Prediction (with observed advertising)" and "Counterfactual (no kynect advertising)".

[c] Difference (%) between "Prediction (with observed advertising)" and "Counterfactual (no insurance company advertising)".

For the latter, the relationship (marginal effect) between the kynect advertising volume during open enrollment and all 3 enrollment outcomes was not statistically significant. This is not surprising given the weaker relationship that we observed in our main results for the enrollment outcomes compared with the information-seeking outcomes, particularly in light of the potential for automatic re-enrollment driving activity when the 2014 open enrollment period is excluded from the analysis.

Table 2 contains estimates of information-seeking and enrollment activity attributable to kynect and insurance company advertising, using our models to predict the average weekly activity given observed advertising and 2 counterfactuals (no kynect advertising, no insurance company advertising). With the substantial drop and eventual elimination of kynect advertising during the 2016 open enrollment period, we used the 2014 and 2015 open enrollment periods as the setting for this portion of the analysis, given the relative stability in state-sponsored advertising and political support for kynect (base scenario). Based on our estimates, kynect advertising was associated with approximately 23% of page views, 22% of visits, and 37% of unique visitors to the kynect website compared with only single digits (6%, 3%, and 6%, respectively) for insurance company advertising. Approximately 38% of Web-based and 28% of total applications were associated with kynect advertising. Insurance company advertising was estimated to be a negative contributor to applications (−30% of Web-based and −11% of total applications), likely because it drove consumers to carriers to enroll in off-exchange plan offerings rather than directing them to the Marketplace [21].

Discussion

Our study is the first to use weekly variation in television advertising to explain consumer engagement with the Marketplace. Specifically, we found significant dose-response effects of the weekly state-sponsored advertising volume during open enrollment on information-seeking behavior (page views, visits, and unique visitors) and enrollment activity (applications). This research offers new evidence of the importance of health insurance advertising in the context of the ACA. Our findings differ from other research on health insurance marketing, which generally find small effects. For example, a recent analysis of the Medicare Advantage market, where television advertising for a large national insurer was found to have a limited effect on encouraging switching from traditional Medicare to Medicare Advantage [22]. This is in contrast to analyses of the first open enrollment period for the ACA in which significant positive effects of television advertising were found [10,11]. Covered California, the state-based Marketplace for the state bearing its name, claims a “more than three-to-one return on investment” on its marketing spending, which “likely lower[ed] premiums by 6 to 8 percent” through improved enrollment and lower risk scores for its enrolled population [12,23].

However, the study is not without limitations. The lack of outcome data at the county or media market level necessitated the construction of statewide population-weighted averages of advertising volume, which reduced variation in our explanatory variables of interest. This is a particular issue owing to border markets where federal and other state advertising spillover into the state are averaged into the statewide advertising measures. Weekly counts of advertising are a cruder measure of exposure than the more desirable gross rating points (which we do not have), which incorporates the ratings for the program during which an ad was aired to provide a more detailed description of the potential reach [18,19]. It is also important to note that Medicaid enrollment is not reflected in our enrollment outcome. Kentucky saw a 6.4 percentage point increase in Medicaid coverage from 2012 to 2015 compared with only a 0.9 percentage point increase in individual nongroup coverage over the same period [24]. The state had approximately half a million enrollees in its Medicaid expansion versus just over 100,000 Marketplace enrollees at its peak in 2015 [24,25]. Although we found no relationship between state-sponsored advertising and enrollment in Marketplace plans, we likely would have observed one if our enrollment measure was more broad (Marketplace plus Medicaid) as it was for the application outcomes that do capture those consumers who may ultimately enroll in Medicaid. Indeed, in a national county-level analysis, a strong relationship was observed between the volume of state-sponsored advertising and Medicaid enrollment [10]. There is no obvious control group because all state-based Marketplaces carry distinct branding and operate in varying political and health insurance market contexts, so we used an interrupted time series approach as the next best option to counteract this weakness to the extent possible, using the time periods outside of open enrollment to establish a baseline level of activity to compare with the open enrollment periods. While dose-response evidence helps us come closer to understanding this relationship, it is not definitive causal evidence. We are measuring two contemporaneous phenomena as including lagged and year-specific effects stretched beyond the limitations of what our data would allow. However, despite these limitations, we were still able to detect strong relationships between the volume of television advertising and our outcomes, reinforcing the role advertising plays in increasing consumer engagement and maximizing potential enrollment. Kentucky ran a highly successful and well-designed campaign in the context of a state-based Marketplace. Therefore, the strength of the dose-response relationship may not generalize to other state-based Marketplaces or those that are federally run due to variation in messaging strategy, demographics of the Marketplace eligible population, Medicaid expansion status, and other factors. This study does not specifically address the message content and tagging (ie, phone numbers, URLs) of health insurance advertising by various ad sponsors, which may be related to differences in their effectiveness in driving Marketplace activity.

Finally, we only assessed television advertising, which took place in the context of larger media campaigns by some (but not all) sponsors. Television is still the dominant news source and advertising medium, despite the growth of digital advertising in recent years [26,27]. It is not feasible from a data availability standpoint to control for all potential advertising channels (eg, radio, digital, print, and out-of-home); however, to the extent that the patterns and targeting of advertising in other channels did not substantially diverge over time from that of the television advertising, this is not an issue for our analysis.

The decision to eliminate television advertising and cut funding for advertising by 90% overall in the federally run Marketplace for the 2018 open enrollment period was justified by administration representatives citing the “diminishing returns” to advertising demonstrated during previous years [1]. However, this has been disputed by members of the Obama administration [28]. The reported shift to a heavier focus on digital advertising may allow for reaching a sizable population at a fraction of the cost but should not be expected to provide the same impact as a television campaign, as TV still garners a majority of viewing time [29]. This conclusion is supported not only by this study but also by evidence from an ongoing national health education campaign [30]. In addition, our results indicate that we should not expect insurance company advertising to help fill the void in terms of generating consumer engagement with the Marketplace without modification of their ad content (eg, explicitly directing consumers to healthcare.gov rather than their own websites). Finally, our findings also speak to another decision related to the fifth open enrollment period under the Trump administration. The large drops in activity that we observed during the week of Thanksgiving suggest that shortening the open enrollment period from 90 to 45 days (from November 1 to December 15) could have further negative consequences, as the Thanksgiving week represents a substantial proportion of the shortened enrollment period.

Approaching the recently completed fifth open enrollment period, 4 in 10 uninsured Americans were still unaware of the Marketplace and a majority of prior Marketplace enrollees were uncertain about the timing of open enrollment [31,32]. Our findings from Kentucky indicate that state-sponsored advertising appears to be highly correlated with traffic to the Marketplace website and applications for coverage—with the kynect television campaign associated with nearly 40% of unique visitors and Web-based applications. This resonates with prior research, albeit limited, that finds a strong link between exposure to Marketplace advertising with perceptions and knowledge about the ACA, improved (lower) risk pools, and reductions in the uninsured rate [8-12]. Insurance company advertising seemed to play a limited role in driving consumers to engage with the Marketplace, likely because their advertising was more geared toward creating brand preference and gaining market share among enrollees (in the Marketplace, but also in employer-sponsored plans) rather than serving as a call to action. These findings raise questions about what role government—as a social institution, rather than a legislative and administrative body—should play in promoting its programs to Americans, particularly for an issue as important and politically charged as health care.

Acknowledgments

This study was supported by two grants from the Robert Wood Johnson Foundation (State Health Access Reform Evaluation [72179; to EFF, LB, and SEG] and Health Policy Research Scholars [73923; to PRS]). The Foundation had no part in the study design, analysis, or decision to publish our findings.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Ad sponsors and sponsor type, Kentucky, October 1, 2013–January 31, 2016.

[PDF File (Adobe PDF File), 35KB - jmir_v20i10e10872_app1.pdf]

Multimedia Appendix 2

Correlation matrix.

[PDF File (Adobe PDF File), 30KB - jmir_v20i10e10872_app2.pdf]

Multimedia Appendix 3

Information-seeking behavior models, Kentucky, October 1, 2013–January 31, 2016.

[PDF File (Adobe PDF File), 91KB - jmir_v20i10e10872_app3.pdf]

Multimedia Appendix 4

Enrollment activity models, Kentucky, October 1, 2013–January 31, 2016.

[PDF File (Adobe PDF File), 85KB - jmir_v20i10e10872_app4.pdf]

References

1. Kliff S. Trump is slashing Obamacare's advertising budget by 90%. 2017 Aug 31. 2017 URL: https://www.vox.com/2017/8/31/16236280/trump-obamacare-outreach-ads [accessed 2017-12-01] [WebCite Cache ID 6ybyo3m6Z]
2. Johnson C. Trump administration ends Affordable Care Act assistance contracts in 18 cities. 2017 URL: https://www.pbs.org/newshour/health/trump-administration-ends-affordable-care-act-assistance-contracts-18-cities [accessed 2018-04-12] [WebCite Cache ID 6ybyvIU1s]
3. Jost TS. ACA Open Enrollment Starts Amidst Tumult. Health Aff (Millwood) 2017 Dec;36(12):2044-2045. [doi: 10.1377/hlthaff.2017.1420] [Medline: 29200338]
4. Pear R. Days Before a Deadline, Trump Team Cancels Ads for Obama Health Plan. 2017 Jan 26. 2017 URL: https://www.nytimes.com/2017/01/26/us/politics/donald-trump-obamacare-ads.html [accessed 2017-07-24] [WebCite Cache ID 6ybz0SToT]
5. The White House. Executive Order Minimizing the Economic Burden of the Patient Protection and Affordable Care Act Pending Repeal. 2017 Jan 24. 2017 URL: https://www.federalregister.gov/documents/2017/01/24/2017-01799/minimizing-the-economic-burden-of-the-patient-protection-and-affordable-care-act-pending-repeal [accessed 2018-09-24] [WebCite Cache ID 72fvgCuC7]
6. US Department of Health and Human Services, Centers for Medicare and Medicaid Services. Federal Register. 2017 Feb 17. Patient Protection and Affordable Care Act; Market Stabilization URL: https://www.gpo.gov/fdsys/search/pagedetails.action?granuleId=2017-03027&packageId=FR-2017-02-17
7. Obama B. United States Health Care Reform: Progress to Date and Next Steps. JAMA 2016 Aug 02;316(5):525-532 [FREE Full text] [doi: 10.1001/jama.2016.9797] [Medline: 27400401]
8. Sommers BD, Maylone B, Nguyen KH, Blendon RJ, Epstein AM. The Impact Of State Policies On ACA Applications And Enrollment Among Low-Income Adults In Arkansas, Kentucky, And Texas. Health Aff (Millwood) 2015 Jun;34(6):1010-1018. [doi: 10.1377/hlthaff.2015.0215] [Medline: 26056207]
9. Fowler EF, Baum LM, Barry CL, Niederdeppe J, Gollust SE. Media Messages and Perceptions of the Affordable Care Act during the Early Phase of Implementation. J Health Polit Policy Law 2017 Feb;42(1):167-195. [doi: 10.1215/03616878-3702806] [Medline: 27729445]
10. Karaca-Mandic P, Wilcock A, Baum L, Barry CL, Fowler EF, Niederdeppe J, et al. The Volume Of TV Advertisements During The ACA's First Enrollment Period Was Associated With Increased Insurance Coverage. Health Aff (Millwood) 2017 Apr 01;36(4):747-754. [doi: 10.1377/hlthaff.2016.1440] [Medline: 28298432]
11. Gollust SE, Wilcock A, Fowler EF, Barry CL, Niederdeppe J, Baum L, et al. TV Advertising Volumes Were Associated With Insurance Marketplace Shopping And Enrollment In 2014. Health Aff (Millwood) 2018 Jun;37(6):956-963. [doi: 10.1377/hlthaff.2017.1507] [Medline: 29863916]
12. Covered California. MARKETING MATTERS: Lessons From California to Promote Stability and Lower Costs in National and State Individual Insurance Markets URL: http://hbex.coveredca.com/data-research/library/CoveredCA_Marketing_Matters_9-17.pdf [accessed 2018-04-12] [WebCite Cache ID 6ybz9o4MA]
13. Gallup. Kentucky, Arkansas Post Largest Drops in Uninsured Rates URL: https://news.gallup.com/poll/203501/kentucky-arkansas-post-largest-drops-uninsured-rates.aspx [accessed 2018-08-02] [WebCite Cache ID 71NmyngaI]

14. Effie. kynect: Kentucky's Healthcare Connection kynect: For Every Kentuckian URL: https://www.effie.org/case_database/case/NA_2015_9537 [accessed 2018-08-02] [WebCite Cache ID 71Nn457ww]
15. The Lane Report. Kynect's technology program coordinator receives national recognition URL: http://www.lanereport.com/55789/2015/10/kynects-technology-program-coordinator-receives-national-recognition/ [accessed 2017-12-01] [WebCite Cache ID 71NnGiyne]
16. Courier Journal. State ends kynect advertising campaign URL: https://www.courier-journal.com/story/news/politics/2015/12/23/state-ends-kynect-advertising-campaign/77769562/ [accessed 2017-03-01] [WebCite Cache ID 71NnNjbxF]
17. Polidata. County-Based Regions and Markets for Kentucky. 2002 URL: http://www.polidata.us/pub/maps/rg2000/ky_reg.pdf [accessed 2018-08-02] [WebCite Cache ID 71NnQtOBJ]
18. Davis KC, Alexander RL, Shafer P, Mann N, Malarcher A, Zhang L. The Dose-Response Relationship Between Tobacco Education Advertising and Calls to Quitlines in the United States, March-June, 2012. Prev Chronic Dis 2015 Nov 05;12:E191 [FREE Full text] [doi: 10.5888/pcd12.150157] [Medline: 26542143]
19. Shafer PR, Davis KC, Patel D, Rodes R, Beistle D. Association Between Media Dose, Ad Tagging, and Changes in Web Traffic for a National Tobacco Education Campaign: A Market-Level Longitudinal Study. J Med Internet Res 2016;18(2):e39 [FREE Full text] [doi: 10.2196/jmir.5343] [Medline: 26887959]
20. StataCorp. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP; 2015.
21. Goddeeris JH, McMorrow S, Kenney GM. Off-Marketplace Enrollment Remains An Important Part Of Health Insurance Under The ACA. Health Aff (Millwood) 2017 Aug 01;36(8):1489-1494. [doi: 10.1377/hlthaff.2017.0085] [Medline: 28784743]
22. Shapiro, B. Advertising in Health Insurance Markets. 2017. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2789904 [WebCite Cache ID 71NnXSrlF]
23. Bingham A, Cohen M, Bertko J. Health Affairs Blog. 2018 Jul 11. National vs. California Comparison: Detailed Data Help Explain The Risk Differences Which Drive Covered California's Success URL: https://www.healthaffairs.org/do/10.1377/hblog20180710.459445/full/ [accessed 2018-07-15] [WebCite Cache ID 71NnaVLPF]
24. State Health Access Data Assistance Center. 2017. Final report: Study of the impact of the ACA implementation in Kentucky URL: https://www.healthy-ky.org/res/images/resources/Impact-of-the-ACA-in-KY_FINAL-Report.pdf [accessed 2017-12-01] [WebCite Cache ID 71NncbX1s]
25. US Department of Health and Human Services. 2015. Health insurance marketplaces open enrollment period: March enrollment report URL: https://aspe.hhs.gov/system/files/pdf/83656/ib_2015mar_enrollment.pdf [accessed 2017-12-01] [WebCite Cache ID 71NnelTtA]
26. Bialik K, Matsa KE. Pew Research Center. 2017 Oct 04. Key trends in social and digital news media URL: http://www.pewresearch.org/fact-tank/2017/10/04/key-trends-in-social-and-digital-news-media/ [accessed 2018-08-02] [WebCite Cache ID 71NnigSoy]
27. Pew Research Center-Journalism & Media. 2018 Jun 06. Digital and non-digital advertising revenue URL: http://www.journalism.org/chart/sotnm-digital-and-non-digital-advertising-revenue/ [accessed 2018-08-02] [WebCite Cache ID 71NnnHZbZ]
28. Lodes L. Vox-Understand the News. 2017 Sep 15. I ran ACA outreach under Obama. Trump's funding cuts could ruin the health care law URL: https://www.vox.com/the-big-idea/2017/9/12/16294784/aca-outreach-advertising-sabotage-funding [accessed 2018-08-03] [WebCite Cache ID 71Nnqy9Sp]
29. Nielsen. 2017. Over 92% of all adult viewing in the US is done on the TV screen URL: http://www.nielsen.com/us/en/insights/news/2017/over-92-percent-of-all-adult-viewing-in-the-us-is-done-on-the-tv-screen.html [accessed 2017-12-01] [WebCite Cache ID 71Nnvb4jC]
30. Davis KC, Shafer PR, Rodes R, Kim A, Hansen H, Patel D, et al. Does Digital Video Advertising Increase Population-Level Reach of Multimedia Campaigns? Evidence From the 2013 Tips From Former Smokers Campaign. J Med Internet Res 2016 Sep 14;18(9):e235 [FREE Full text] [doi: 10.2196/jmir.5683] [Medline: 27627853]
31. Kirzinger A, Hamel L, DiJulio B, Muñana C, Brodie M. Kaiser Family Foundation. 2017 Oct. Kaiser Health Tracking Poll – October 2017: Experiences of the Non-Group Marketplace Enrollees URL: https://www.kff.org/health-reform/poll-finding/kaiser-health-tracking-poll-october-2017-experiences-of-the-non-group-marketplace-enrollees/ [accessed 2018-08-03] [WebCite Cache ID 71No0qRRS]
32. Collins SR, Gunja MZ, Doty MM. The Commonwealth Fund. 2017 Sep 07. Following the ACA Repeal-and-Replace Effort, Where Does the US Stand on Insurance Coverage? Findings from the Commonwealth Fund Affordable Care Act Tracking Survey, March–June 2017 URL: https://www.commonwealthfund.org/publications/issue-briefs/2017/sep/following-aca-repeal-and-replace-effort-where-does-us-stand?redirect_source=/publications/issue-briefs/2017/sep/post-aca-repeal-and-replace-health-insurance-coverage [accessed 2018-08-03] [WebCite Cache ID 71No606J1]

Abbreviations

ACA: Affordable Care Act
CMAG: Campaign Media Analysis Group

KHBE: Kentucky Health Benefit Exchange

Edited by G Eysenbach; submitted 24.04.18; peer-reviewed by S Zwier, J Villegas; comments to author 24.05.18; revised version received 19.07.18; accepted 01.08.18; published 25.10.18.

Please cite as:
Shafer PR, Fowler EF, Baum L, Gollust SE
Television Advertising and Health Insurance Marketplace Consumer Engagement in Kentucky: A Natural Experiment
J Med Internet Res 2018;20(10):e10872
URL: http://www.jmir.org/2018/10/e10872/
doi:10.2196/10872
PMID:30361198

Review

Telerehabilitation for Stroke Survivors: Systematic Review and Meta-Analysis

Huidi Tchero[1], MD; Maturin Tabue Teguo[2,3,4,5], MD, PhD; Annie Lannuzel[6], MD, PhD; Emmanuel Rusch[7], MD, PhD

[1]Unit of Wounds and Healing, Department of Trauma and Orthopaedic Surgery, Centre Hospitalier Louis Constant Fleming Saint Martin, Saint Martin, Guadeloupe, France

[2]Centre Hospitalier Universitaire de Guadeloupe, Université des Antilles, Guadeloupe, France

[3]Le Centre de Recherche Inserm - U1219, Université de Bordeaux, Bordeaux, France

[4]Institut de santé publique, d'épidémiologie et de développement, Université de Bordeaux, Bordeaux, France

[5]Le Gérontopôle du Centre Hospitalier Universitaire de Toulouse, Toulouse, France

[6]Department of Neurology, University Hospital Center Pointe-à-Pitre, Guadeloupe, France

[7]Service d'Information Médicale, Epidémiologie et Economie de la Santé, Centre Hospitalier Régional Universitaire de Tours, Tours, France

Corresponding Author:
Huidi Tchero, MD
Unit of Wounds and Healing
Department of Trauma and Orthopaedic Surgery
Centre Hospitalier Louis Constant Fleming Saint Martin
BP 381 - Spring Concordia
Saint Martin, Guadeloupe, 97150
France
Phone: 33 0690660959
Email: h.tchero@chsaintmartin.fr

Abstract

Background: Telerehabilitation is an emerging technology through which medical rehabilitation care can be provided from a distance.

Objective: This systematic review and meta-analysis aims to investigate the efficacy of telerehabilitation in poststroke patients.

Methods: Eligible randomized controlled trials (RCTs) were identified by searching MEDLINE, Cochrane Central, and Web of Science databases. Continuous data were extracted for relevant outcomes and analyzed using the RevMan software as the standardized mean difference (SMD) and 95% CI in a fixed-effect meta-analysis model.

Results: We included 15 studies (1339 patients) in our systematic review, while only 12 were included in the pooled analysis. The combined effect estimate showed no significant differences between the telerehabilitation and control groups in terms of the Barthel Index (SMD –0.05, 95% CI –0.18 to 0.08), Berg Balance Scale (SMD –0.04, 95% CI –0.34 to 0.26), Fugl-Meyer Upper Extremity (SMD 0.50, 95% CI –0.09 to 1.09), and Stroke Impact Scale (mobility subscale; SMD 0.18, 95% CI –0.13 to 0.48]) scores. Moreover, the majority of included studies showed that both groups were comparable in terms of health-related quality of life (of stroke survivors), Caregiver Strain Index, and patients' satisfaction with care. One study showed that the cost of telerehabilitation was lower than usual care by US $867.

Conclusions: Telerehabilitation can be a suitable alternative to usual rehabilitation care in poststroke patients, especially in remote or underserved areas. Larger studies are needed to evaluate the health-related quality of life and cost-effectiveness with the ongoing improvements in telerehabilitation networks.

(J Med Internet Res 2018;20(10):e10867) doi:10.2196/10867

KEYWORDS

meta-analysis; mobile phone; rehabilitation; stroke; telemedicine

Introduction

Telemedicine is the exchange of medical information from one location to another using electronic communication to achieve clinical health care from a distance [1]. These technologies allow communication between medical staff and patients, as well as the transmission of imaging and other health information data from one place to another [2]. It can be used to accelerate medical emergency services in conditions with narrow therapeutic windows, such as stroke [3] and myocardial infarction [4], and facilitate access to medical services that would not often be available in rural communities [2].

Stroke rehabilitation therapy aims to improve the patients' motor function, health-related quality of life (HRQoL), and psychological well-being [5]. Successful rehabilitation depends on stroke severity, rehabilitation team skills, and the cooperation of patients and their families [6]. However, many patients have reduced access to care because of limited regional and logistic resources; these patient groups could benefit from a system that allows a health professional to provide rehabilitation services from a remote location [7]. A physical medicine or rehabilitation specialist at a hospital can observe patients as they execute movements and monitor their improvement. Quantitative data, such as range of motion and physical force, can be recorded and transported through the network to the hospital for review [8].

Over the past decade, a number of randomized controlled trials (RCTs) have investigated the benefits of telerehabilitation in poststroke patients in comparison to usual rehabilitation methods. These studies showed that telerehabilitation was either equal [9,10] or superior [11,12] to usual rehabilitation in terms of improvements in the activities of daily living and psychological status of patients and their caregivers. This study aims to investigate the benefits of telemedicine in poststroke rehabilitation in a meta-analysis framework.

Methods

This study was performed and reported following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses checklist for systematic reviews of intervention (Multimedia Appendix 1).

Literature Search and Study Selection

On January 14, 2018, we performed a comprehensive search of the following databases: MEDLINE (via PubMed), Cochrane Central, and Web of Science. The following keywords were used with different combinations: "Telemedicine," "Telestroke," "Telerehabilitation," "Stroke," and "Brain infarction" with no filters applied (either by language or period of publication). Multimedia Appendix 2 provides the search strategies of the 3 used databases. In addition, we searched the clinical trial register "Clinicaltrials.gov" for any unpublished or ongoing studies. Furthermore, a manual screening of the bibliography of included studies was performed for any studies we missed during the electronic search.

Studies were considered eligible for inclusion if they fulfilled the following criteria: (1) assessed the efficacy of different telerehabilitation models in poststroke patients and (2) employed an RCT design. We excluded studies that had a nonrandomized or single-arm design or that examined the technical components of the telerehabilitation systems. Two reviewers screened the search results using the criteria mentioned above in 2 subsequent steps: title and abstract screening, followed by full-text screening. When the judgments of both reviewers were not similar, another reviewer solved the discrepancy.

Data Extraction and Outcomes

Two independent reviewers used a preformatted Excel sheet to extract data for the prespecified outcomes, including (1) activities of daily living: Barthel Index and Berg Balance Scale; (2) motor function: Action Research Arm Test (ARAT), Fugl-Meyer Upper Extremity (FM-UE), and Stroke Impact Scale (mobility subscale); (3) HRQoL outcomes and satisfaction with care; and (4) cost-effectiveness. Data were extracted as the mean (SD) of change before and after treatment and then was compared between groups. When these values were not given in the included studies, they were calculated using the equations in the Cochrane Handbook for Systematic reviews of intervention [13]. When numerical data were not available for these outcomes or could not be reliably extracted, they were analyzed in a qualitative approach.

During the extraction, we also evaluated the risk of different forms of bias in the included studies. We used the Cochrane risk of bias tool [13], which deals with the following sources of bias: (1) selection bias (random sequence generation and allocation concealment); (2) performance bias (blinding of participants and outcome assessors); (3) attrition bias (incomplete outcome data); (4) reporting bias (selective reporting); and (5) other sources of bias.

Statistical Analysis and Outcome Interpretation

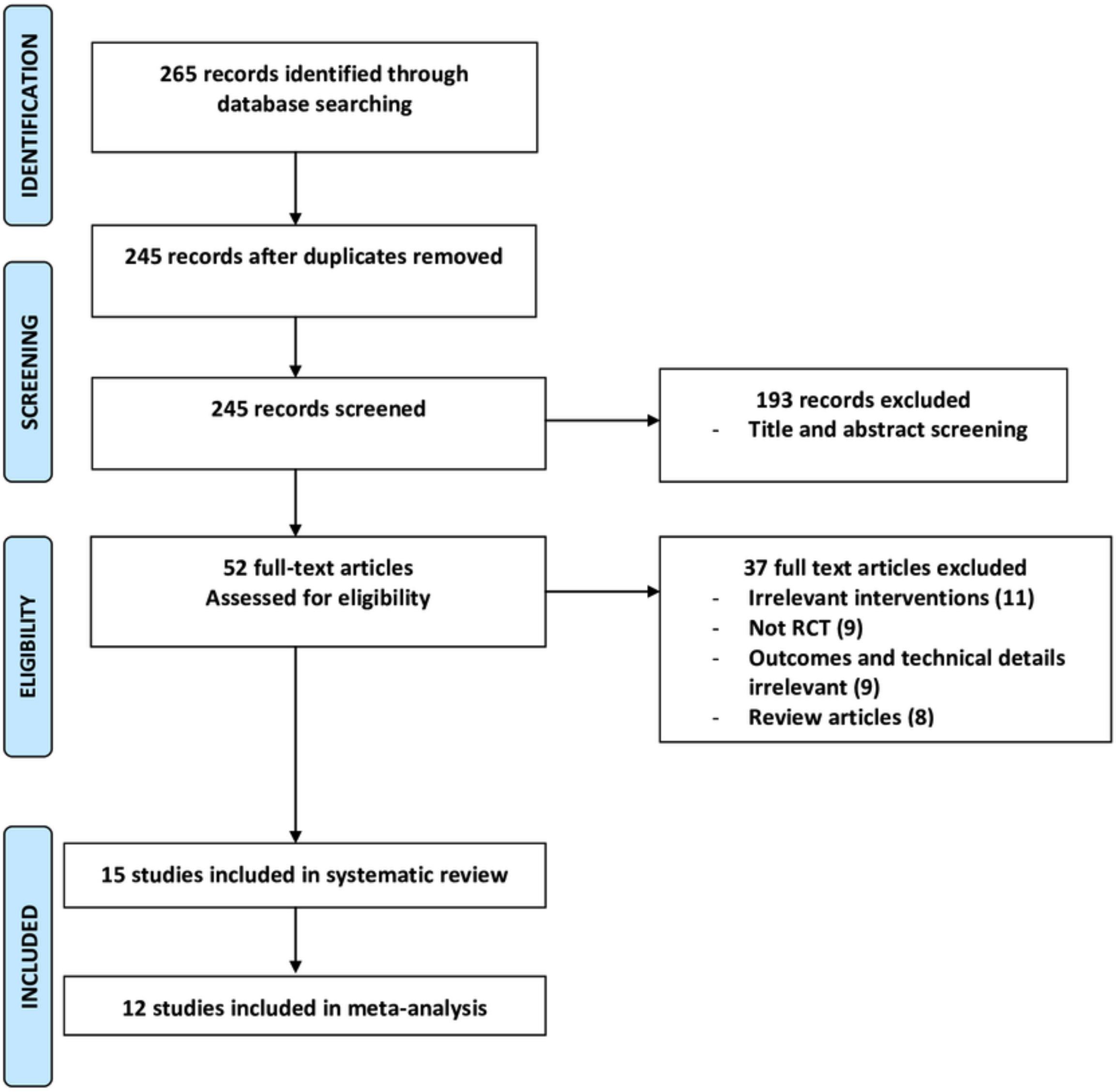
We used the RevMan software (Version 5.3; Cochrane Collaboration, Oxford, England) to perform all statistical analyses in this study. Based on the nature of extracted data (continuous), they were pooled as the standardized mean difference (SMD) with 95% CI, using the inverse variance meta-analysis method. $P<.05$ was considered significant for the effect estimate. The analysis was done first under the fixed-effect model for assuming homogeneity; in case of heterogeneity, we shifted to the random-effects model. Heterogeneity was assessed using the chi-square test ($P<.10$ was considered significant for between-study heterogeneity), and its extent was measured using the I^2 test.

Results

Results of Literature Search

Our literature search retrieved 245 unique records, which were reduced to 52 records after the title and abstract screening. After a meticulous full-text screening, 15 studies (1339 patients) were identified as eligible for our systematic review [9-12,14-24], while only 12 [9-11,16-24] were included in our meta-analysis (1246 patients). Figure 1 shows the details of our screening process.

Figure 1. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram of the study selection process. RCT: randomized controlled trial.



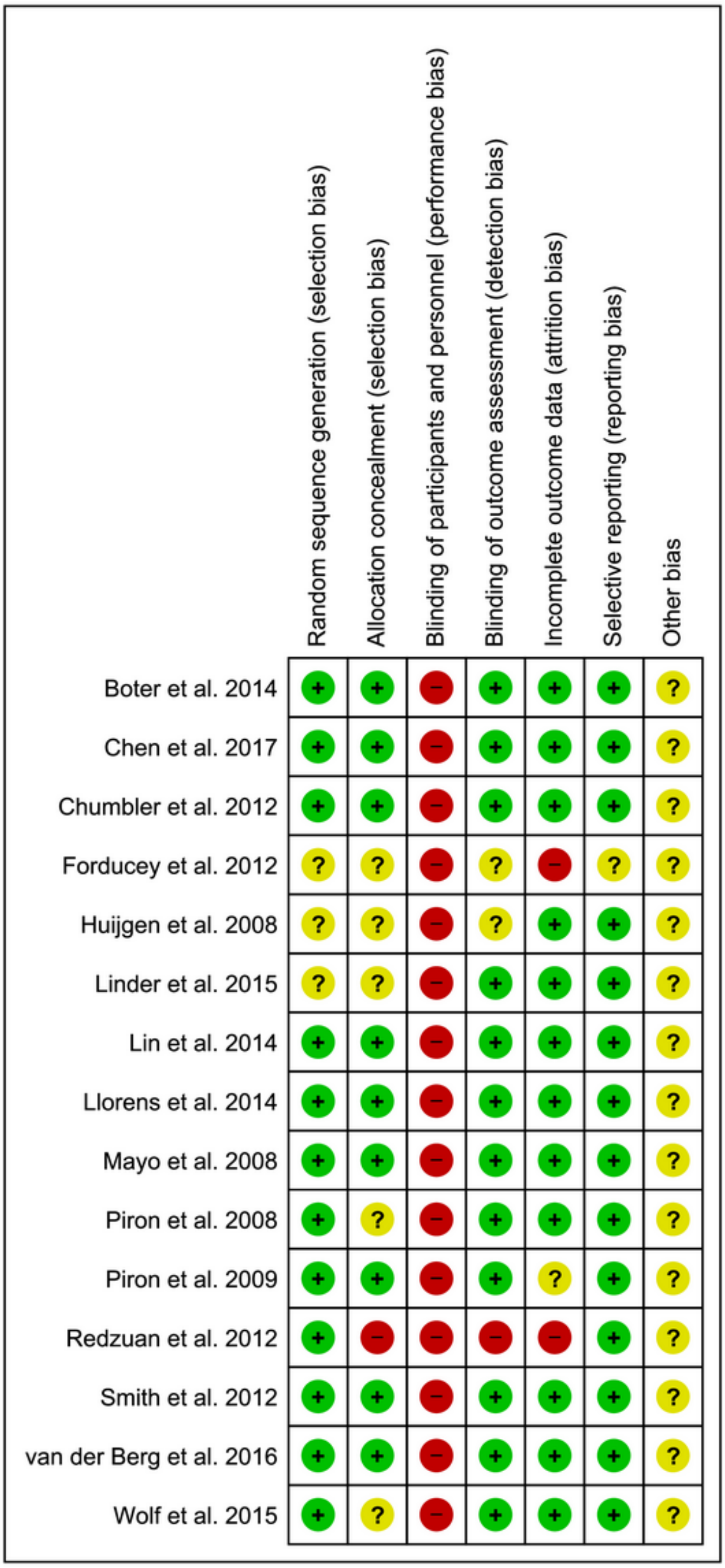
Characteristics of Included Studies

The included RCTs had a sample size ranging between 9 and 536 patients. They compared different models of telerehabilitation to standard rehabilitation care or a home-based exercise program. The follow-up period in these studies ranged between 4 and 24 weeks. Multimedia Appendix 3 summarizes the design of included studies, components of the used telerehabilitation systems, and the baseline characteristics of enrolled patients.

Risk of Bias Assessment Results

All included studies reported adequately on their methods of random sequence generation, blinding of outcome assessors, and reducing the risk of attrition bias, except for 3 trials in each domain. Owing to the nature of the intervention, blinding participants was not possible in all included trials. Only 9 studies reported adequately on their methods of allocation concealment [9-12,14,17,19-21]. Figure 2 summarizes the risk of bias assessment results, with red, green, and yellow colors indicating high, low, and unclear risk of bias, respectively.

Figure 2. Risk of bias assessment summary according to the Cochrane risk of bias tool: Red, green, and yellow colours indicates high, low, and unclear risk of bias, respectively.



	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants and personnel (performance bias)	Blinding of outcome assessment (detection bias)	Incomplete outcome data (attrition bias)	Selective reporting (reporting bias)	Other bias
Boter et al. 2014	+	+	−	+	+	+	?
Chen et al. 2017	+	+	−	+	+	+	?
Chumbler et al. 2012	+	+	−	+	+	+	?
Forducey et al. 2012	?	?	−	?	−	?	?
Huijgen et al. 2008	?	?	−	?	+	+	?
Linder et al. 2015	?	?	−	+	+	+	?
Lin et al. 2014	+	+	−	+	+	+	?
Llorens et al. 2014	+	+	−	+	+	+	?
Mayo et al. 2008	+	+	−	+	+	+	?
Piron et al. 2008	+	?	−	+	+	+	?
Piron et al. 2009	+	+	−	+	?	+	?
Redzuan et al. 2012	+	−	−	−	−	+	?
Smith et al. 2012	+	+	−	+	+	+	?
van der Berg et al. 2016	+	+	−	+	+	+	?
Wolf et al. 2015	+	?	−	+	+	+	?

Results of Outcome Assessment

Activities of Daily Living and Balance Function

Barthel Index

Under the fixed-effect model, pooling data from 6 trials [9-11,17,20,23] showed no significant difference between the telerehabilitation and control groups in terms of the Barthel Index score (SMD –0.05, 95% CI –0.18 to 0.08, *P*=.47, 909 patients). Pooled studies were homogenous (*P*=.51, I^2=0%; Figure 3).

Berg Balance Scale

Under the fixed-effect model, the pooled analysis of data from 4 studies [10,11,17,19] showed no significant difference between the telerehabilitation and control groups in terms of the Berg Balance Scale (SMD –0.04, 95% CI –0.34 to 0.26, *P*=.78, 171 patients). Pooled studies were homogenous (*P*=.77, I^2=0%; Figure 3).

Motor Function

In this study, different scales were used to assess this outcome.

Fugl-Meyer Upper Extremity

Two homogenous studies (*P*=.43, I^2=0%) reported data on the mean FM-UE score in both groups [21,22]. Under the fixed-effect model, the pooled effect estimate showed no significant difference (SMD 0.50, 95% CI –0.09 to 1.09, *P*=.10, 46 patients) between the telerehabilitation and control groups with regard to FM-UE (Figure 4).

Action Research Arm Test

Two homogenous studies (*P*=.93, I^2=0%) provided data on the mean ARAT score in both groups [16,24]; therefore, the analysis was conducted under the fixed-effect model. No significant difference was noted between both groups in terms of the ARAT score between the telerehabilitation and control groups (SMD –0.06, 95% CI –0.46 to 0.33, *P*=.75, 98 patients; Figure 4).

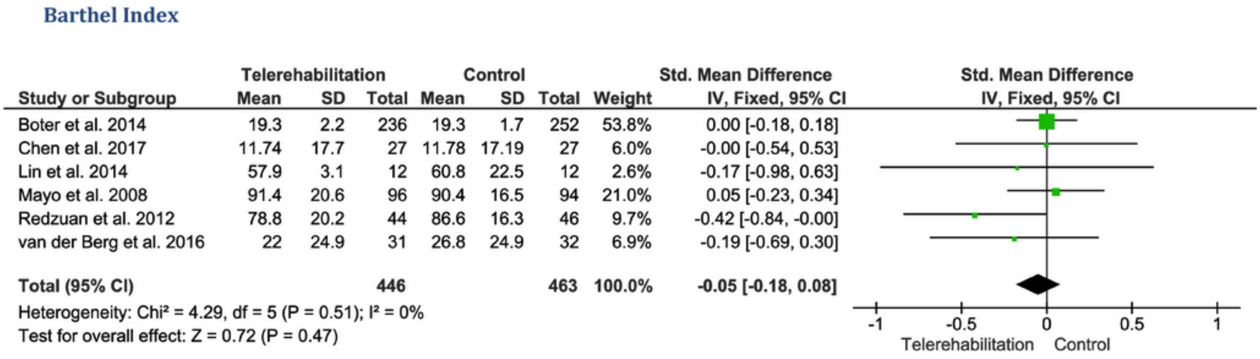
Stroke Impact Scale—Mobility Subscale

Under the fixed-effect model, the pooled effect estimate of 2 studies [11,18] showed no significant difference (SMD 0.18, 95% CI –0.13 to 0.48, *P*=.26, 162 patients) between the telerehabilitation and control groups in terms of the Stroke Impact Scale-mobility subscale score. Pooled studies were homogenous (*P*=.87, I^2=0%; Figure 4).

Patients' Quality of Life

Six studies reported on the HRQoL of poststroke patients. Boter reported that telerehabilitation patients achieved better scores on the Short-Form (SF-36) emotional role limitation (mean difference=7.9, 95% CI 0.1 to 15.7) than the control group [9]. However, Forducey et al and Mayo et al showed no significant differences (*P*>.05) between both groups with regard to the SF-12 and the physical component score of the SF-36, respectively [15,20].

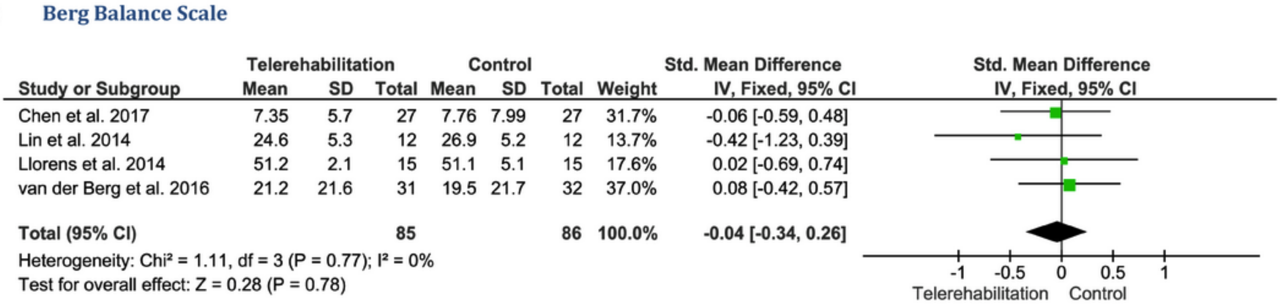
Figure 3. The pooled standardized mean difference between the telerehabilitation and control groups in terms of Barthel Index and Berg Balance Scale scores.



Barthel Index

Study or Subgroup	Telerehabilitation Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Fixed, 95% CI
Boter et al. 2014	19.3	2.2	236	19.3	1.7	252	53.8%	0.00 [-0.18, 0.18]
Chen et al. 2017	11.74	17.7	27	11.78	17.19	27	6.0%	-0.00 [-0.54, 0.53]
Lin et al. 2014	57.9	3.1	12	60.8	22.5	12	2.6%	-0.17 [-0.98, 0.63]
Mayo et al. 2008	91.4	20.6	96	90.4	16.5	94	21.0%	0.05 [-0.23, 0.34]
Redzuan et al. 2012	78.8	20.2	44	86.6	16.3	46	9.7%	-0.42 [-0.84, -0.00]
van der Berg et al. 2016	22	24.9	31	26.8	24.9	32	6.9%	-0.19 [-0.69, 0.30]
Total (95% CI)			**446**			**463**	**100.0%**	**-0.05 [-0.18, 0.08]**

Heterogeneity: Chi² = 4.29, df = 5 (P = 0.51); I² = 0%
Test for overall effect: Z = 0.72 (P = 0.47)

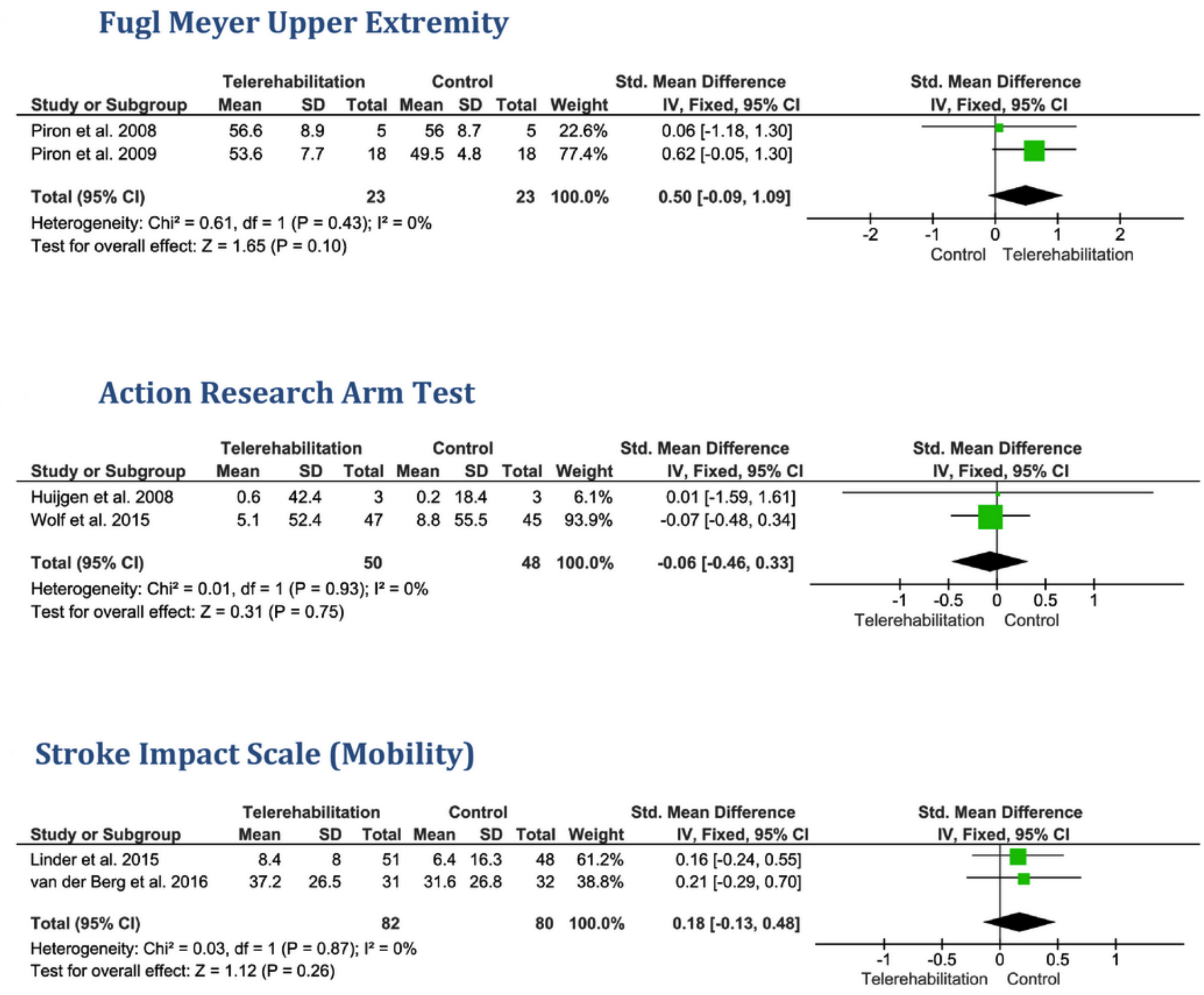


Berg Balance Scale

Study or Subgroup	Telerehabilitation Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Fixed, 95% CI
Chen et al. 2017	7.35	5.7	27	7.76	7.99	27	31.7%	-0.06 [-0.59, 0.48]
Lin et al. 2014	24.6	5.3	12	26.9	5.2	12	13.7%	-0.42 [-1.23, 0.39]
Llorens et al. 2014	51.2	2.1	15	51.1	5.1	15	17.6%	0.02 [-0.69, 0.74]
van der Berg et al. 2016	21.2	21.6	31	19.5	21.7	32	37.0%	0.08 [-0.42, 0.57]
Total (95% CI)			**85**			**86**	**100.0%**	**-0.04 [-0.34, 0.26]**

Heterogeneity: Chi² = 1.11, df = 3 (P = 0.77); I² = 0%
Test for overall effect: Z = 0.28 (P = 0.78)

Figure 4. The pooled standardized mean difference between the telerehabilitation and control groups in terms of Fugl-Meyer Upper Extremity, Action Research Arm Test, and Stroke Impact Scale (mobility) scores.



Fugl Meyer Upper Extremity

Study or Subgroup	Telerehabilitation Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Fixed, 95% CI
Piron et al. 2008	56.6	8.9	5	56	8.7	5	22.6%	0.06 [-1.18, 1.30]
Piron et al. 2009	53.6	7.7	18	49.5	4.8	18	77.4%	0.62 [-0.05, 1.30]
Total (95% CI)			**23**			**23**	**100.0%**	**0.50 [-0.09, 1.09]**

Heterogeneity: Chi² = 0.61, df = 1 (P = 0.43); I² = 0%

Test for overall effect: Z = 1.65 (P = 0.10)

Action Research Arm Test

Study or Subgroup	Telerehabilitation Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Fixed, 95% CI
Huijgen et al. 2008	0.6	42.4	3	0.2	18.4	3	6.1%	0.01 [-1.59, 1.61]
Wolf et al. 2015	5.1	52.4	47	8.8	55.5	45	93.9%	-0.07 [-0.48, 0.34]
Total (95% CI)			**50**			**48**	**100.0%**	**-0.06 [-0.46, 0.33]**

Heterogeneity: Chi² = 0.01, df = 1 (P = 0.93); I² = 0%

Test for overall effect: Z = 0.31 (P = 0.75)

Stroke Impact Scale (Mobility)

Study or Subgroup	Telerehabilitation Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Fixed, 95% CI
Linder et al. 2015	8.4	8	51	6.4	16.3	48	61.2%	0.16 [-0.24, 0.55]
van der Berg et al. 2016	37.2	26.5	31	31.6	26.8	32	38.8%	0.21 [-0.29, 0.70]
Total (95% CI)			**82**			**80**	**100.0%**	**0.18 [-0.13, 0.48]**

Heterogeneity: Chi² = 0.03, df = 1 (P = 0.87); I² = 0%

Test for overall effect: Z = 1.12 (P = 0.26)

Two studies used 2 different version of the Functional Independence Measure—self-administered and telephone versions. Both studies recorded no differences between the telerehabilitation and control groups [14,15]. Interestingly, Smith et al compared the effects of both treatments on the outcomes of mastery, self-esteem, and social support and reported no significant effects of either treatment on these outcomes [12].

In addition, 4 of 6 studies assessed the treatment effects on depression. Two studies by Linder et al and Smith et al recorded no differences between both rehabilitation methods (telerehabilitation and usual rehabilitation) on the Center for Epidemiological Studies-Depression scale in poststroke patients [12,18]. Similarly, Boter et al and Mayo et al reported no significant difference between both groups on the Hospital Depression Scale and the Geriatric Depression Scale [9,20]. Only one study by Redzuan et al reported comparable rates of poststroke complications in these groups but did not specify the nature of these complications [23].

Caregivers' Quality of Life

Five included studies reported on the quality of life in caregivers of poststroke patients. Of these, 4 studies showed no significant difference (*P*<.05) between the 2 rehabilitation modalities (telerehabilitation and usual rehabilitation) in terms of the Caregiver Strain Index [9-11,23]. On the other hand, Smith et al reported that caregivers in the telerehabilitation group had lower depression scores than those in the usual care group [12].

Satisfaction With Care

Three studies reported on the patient satisfaction with care in both groups. Of these, 2 these studies showed no significant difference between the telerehabilitation and usual care groups in terms of patient satisfaction scores [9,17]. While Piron et al randomized 10 patients into a virtual reality virtual reality-telerehabilitation group and virtual reality hospital-based intervention. Using a modified satisfaction questionnaire, they reported that tele-virtual reality patients achieved equal or higher scores to hospital-based virtual realityR patients in almost all points; however, a significant difference in motor performance was only noted in the tele-virtual reality group [22].

Cost-Effectiveness

Only one study of high methodological quality (according to the Cochrane risk of bias tool) by Lloréns et al reported data on the cost-effectiveness outcome. They calculated that for 1 participant, the cost of telerehabilitation was lower than that of usual care by about US $654 (US $1490.23 and US $853.61 for in-clinic and home rehabilitation programs, respectively) [19]. Although setting the virtual reality system at home required US $800, the telerehabilitation arm required fewer work hours by physical therapists and it eliminated the cost of round trips to the clinic with every session.

Discussion

This systematic review showed that patients on telerehabilitation achieve comparable restoration of daily of life activities and HRQoL to those on usual care rehabilitation. Moreover, caregivers of stroke survivors in both groups had a comparable quality of life (as assessed by the Caregiver Strain Index), and one study [12] reported lower rates of depression in the telerehabilitation group. Satisfaction with care remains a problem in poststroke rehabilitation as most included studies showed that telerehabilitation failed to improve the patients' satisfaction with care.

The comparable improvement in motor performance in the telerehabilitation and usual care groups was evident on all motor assessment scales; this adds to the reliability of our findings that telerehabilitation can produce significant motor improvements. Moreover, a previous meta-analysis showed that when telemedicine is used to deliver thrombolytic therapy for stroke patients, it produced comparable rates of favorable clinical outcomes at 90 days as those produced by in-hospital thrombolysis [3]. Therefore, telemedicine can assist with improving the motor function from the onset of stroke, not only during rehabilitation. The improved motor performance would further translate into improved activities of daily living and HRQoL.

Another interesting finding of our review is that telerehabilitation produced comparable improvement of the patients' HRQoL to usual rehabilitation; this extended beyond stroke survivors to include their caregivers as well. This is essential because several studies have shown high rates of depression and quality of life impairment among caregivers of stroke survivors, which negatively influences their supportive functions [25,26]. However, research on the caregiver's quality of life and interventions to improve their performance is not adequate. Therefore, confirming the value of telemedicine in this regard should be a focus of future studies.

Compared with usual rehabilitation, telerehabilitation offers several advantages, including easier access, mentoring for disabled stroke patients, and the ability of patients to self-record on their pain, mood, and activity [27]. Unfortunately, several barriers limit the spreading of telerehabilitation; these barriers include administrative licensing, medicolegal ambiguity, and financial sustainability [28]. Another barrier, especially in low-income countries (where telerehabilitation would be most needed), is the lack of technological infrastructure. A cross-sectional study (on 100 stroke survivors) in a Ghanaian outpatient neurology clinic demonstrated that 80%-93% of patients had a positive attitude toward telerehabilitation interventions; however, only 35% of them had smartphones [29]. Further development of telerehabilitation networks is essential to overcome these barriers [30].

The included studies used different models of telerehabilitation. For example, some studies used only telephone calls, while others used videoconferencing, educational videos, Web-based chats, and virtual reality systems (as illustrated in Multimedia Appendix 3). Moreover, the duration of rehabilitation programs and frequency of follow-up visits or contact with medical staff differed from a study to another. So far, there are no adequate data in the literature about which model or telerehabilitation tool is optimal for these patients and thus future head-to-head comparative studies are advised.

Regarding the cost-effectiveness, the included study by Llorens et al showed lower cost (by US $654) for the telerehabilitation program than the in-clinic program with similar efficacy. This is in agreement with several previous studies in other patient populations that showed that telerehabilitation could reduce the cost of disease management and patient rehabilitation. The reduction in cost is mostly due to avoiding travel (especially from remote areas where telemedicine is most useful) [31,32]. However, the evidence is conflicting on this outcome, and further studies are needed to confirm it [33,34].

Our systematic review has some strengths. We performed a comprehensive literature search and reported our methodology according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses checklist and the Cochrane Handbook for Systematic Reviews of Interventions. Compared with a former systematic review of 10 studies by Laver et al [35], we investigated the effects of telemedicine on several outcomes—activities of daily living, motor performance, HRQoL, satisfaction with care, and cost-effectiveness. Moreover, we evaluated the benefits of telemedicine not only on stroke survivors but also on the caregiver's quality of life.

However, our meta-analysis has some limitations. First, the relatively small sample size in included studies restricts the generalizability of our findings. Second, some outcomes could not be analyzed quantitatively because of the heterogeneity of data in the included studies (different scales of measurements or data formats). Moreover, we could not assess the risk of publication bias because according to Egger et al, funnel plot-based methods are not accurate for <10 included studies per outcome [36].

Hence, larger RCTs are required to confirm the current evidence and provide more data on outcomes such as HRQoL and cost-effectiveness. Data reporting should be performed in a clear standardized format to enable reliable extraction for future meta-analysis studies. Our search of clinicaltrials.gov registry retrieved 19 ongoing studies that are active or still recruiting participants such as NCT02665052, NCT02360488, and NCT01157195. The results of these studies are eagerly awaited. Moreover, it would be interesting for clinicians to investigate the benefits of using telerehabilitation to supplement the usual care. Of note, these trials should not necessarily show that telerehabilitation achieves high outcomes, but confirmation of comparable outcomes is needed.

In conclusion, telerehabilitation can be a suitable alternative to usual rehabilitation care in poststroke patients. This may have potential implications for patients, especially in remote or underserved areas. Nevertheless, larger studies are needed to evaluate the quality of life and cost-effectiveness with the ongoing advances in telerehabilitation systems.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) checklist.
[PDF File (Adobe PDF File), 42 KB - jmir_v20i10e10867_app1.pdf]

Multimedia Appendix 2
Literature search strategy.
[PDF File (Adobe PDF File), 22 KB - jmir_v20i10e10867_app2.pdf]

Multimedia Appendix 3
Risk of bias assessment in included randomized controlled trials.
[PDF File (Adobe PDF File), 58 KB - jmir_v20i10e10867_app3.pdf]

References

1. Smith AC, Bensink M, Armfield N, Stillman J, Caffery L. Telemedicine and rural health care applications. J Postgrad Med 2005;51(4):286-293 [FREE Full text] [Medline: 16388171]
2. Matusitz J, Breen GM. Telemedicine: its effects on health communication. Health Commun 2007;21(1):73-83. [doi: 10.1080/10410230701283439] [Medline: 17461754]
3. Baratloo A, Rahimpour L, Abushouk A, Safari S, Lee C, Abdalvand A. Effects of Telestroke on Thrombolysis Times and Outcomes: A Meta-analysis. Prehosp Emerg Care 2018;22(4):472-484. [doi: 10.1080/10903127.2017.1408728] [Medline: 29345529]
4. de Waure C, Cadeddu C, Gualano M, Ricciardi W. Telemedicine for the reduction of myocardial infarction mortality: a systematic review and a meta-analysis of published studies. Telemed J E Health 2012 Jun;18(5):323-328. [doi: 10.1089/tmj.2011.0158] [Medline: 22468983]
5. Hz N, Brady B, Mcgahan L, Teasell R, Skidmore B, Tj D. Canadian Coordinating Office for Health Technology Assessment. 2003. Stroke rehabilitation services: systematic reviews of the clinical and economic evidence URL: http://www.crd.york.ac.uk/CRDWeb/ShowRecord.asp?ID=32003000492 [accessed 2018-09-14] [WebCite Cache ID 72Qez1C4O]
6. de Bustos EM, Vuillier F, Chavot D, Moulin T. Telemedicine in stroke: organizing a network–rationale and baseline principles. Cerebrovasc Dis 2009;27 Suppl 4:1-8. [doi: 10.1159/000213052] [Medline: 19546535]
7. Holden M, Dyar T, Dayan-Cimadoro L. Telerehabilitation using a virtual environment improves upper extremity function in patients with stroke. IEEE Trans Neural Syst Rehabil Eng 2007 Mar;15(1):36-42. [doi: 10.1109/TNSRE.2007.891388] [Medline: 17436874]
8. Cikajlo I, Rudolf M, Goljar N, Burger H, Matjačić Z. Telerehabilitation using virtual reality task can improve balance in patients with stroke. Disabil Rehabil 2012;34(1):13-18. [doi: 10.3109/09638288.2011.583308] [Medline: 21864205]
9. Boter H, HESTIA Study Group. Multicenter randomized controlled trial of an outreach nursing support program for recently discharged stroke patients. Stroke 2004 Dec;35(12):2867-2872. [doi: 10.1161/01.STR.0000147717.57531.e5] [Medline: 15514186]
10. Chen J, Jin W, Dong WS, Jin Y, Qiao FL, Zhou YF, et al. Effects of Home-based Telesupervising Rehabilitation on Physical Function for Stroke Survivors with Hemiplegia: A Randomized Controlled Trial. Am J Phys Med Rehabil 2017 Mar;96(3):152-160. [doi: 10.1097/PHM.0000000000000559] [Medline: 27386808]
11. van den Berg M, Crotty M, Liu E, Killington M, Kwakkel G, van Wegen E. Early Supported Discharge by Caregiver-Mediated Exercises and e-Health Support After Stroke: A Proof-of-Concept Trial. Stroke 2016 Dec;47(7):1885-1892. [doi: 10.1161/STROKEAHA.116.013431] [Medline: 27301941]
12. Smith G, Egbert N, Dellman-Jenkins M, Nanna K, Palmieri P. Reducing depression in stroke survivors and their informal caregivers: a randomized clinical trial of a Web-based intervention. Rehabil Psychol 2012 Aug;57(3):196-206 [FREE Full text] [doi: 10.1037/a0029587] [Medline: 22946607]
13. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of interventions. In: Cochrane Handbook for Systematic Reviews of Interventions. US: Wiley, John & Sons, Incorporated; 2008.
14. Chumbler N, Quigley P, Li X, Morey M, Rose D, Sanford J. Effects of telerehabilitation on physical function and disability for stroke patients: a randomized, controlled trial. Stroke 2012;43(8):74.

15. Forducey PG, Glueckauf RL, Bergquist TF, Maheu MM, Yutsis M. Telehealth for persons with severe functional disabilities and their caregivers: facilitating self-care management in the home setting. Psychol Serv 2012 May;9(2):144-162 [FREE Full text] [doi: 10.1037/a0028112] [Medline: 22662729]
16. Huijgen B, Vollenbroek-Hutten M, Zampolini M, Opisso E, Bernabeu M, Van NJ. Feasibility of a home-based telerehabilitation system compared to usual care: arm/hand function in patients with stroke, traumatic brain injury and multiple sclerosis. Journal of telemedicine and telecare 2008;14(5):56. [Medline: 18633000]
17. Lin K, Chen C, Chen Y, Huang W, Lai J, Yu S. Bidirectional and multi-user telerehabilitation system: clinical effect on balance, functional activity, and satisfaction in patients with chronic stroke living in long-term care facilities. Sensors 2014;14(7):66.
18. Linder S, Rosenfeldt A, Bay R, Sahu K, Wolf S, Alberts J. Improving Quality of Life and Depression After Stroke Through Telerehabilitation. Am J Occup Ther 2015;69(2):6902290020p1-690229002010 [FREE Full text] [doi: 10.5014/ajot.2015.014498] [Medline: 26122686]
19. Lloréns R, Noé E, Colomer C, Alcañiz M. Effectiveness, usability, and cost-benefit of a virtual reality-based telerehabilitation program for balance recovery after stroke: a randomized controlled trial. Arch Phys Med Rehabil 2015 Mar;96(3):418-425.e2. [doi: 10.1016/j.apmr.2014.10.019] [Medline: 25448245]
20. Mayo N, Nadeau L, Ahmed S, White C, Grad R, Huang A. Bridging the gap: the effectiveness of teaming a stroke coordinator with patient's personal physician on the outcome of stroke. Age and ageing 2007;37(1):8.
21. Piron L, Turolla A, Agostini M, Zucconi C, Cortese F, Zampolini M. Exercises for paretic upper limb after stroke: a combined virtual-reality and telemedicine approach. Journal of Rehabilitation Medicine 2009;41(12):20.
22. Piron L, Turolla A, Tonin P, Piccione F, Lain L, Dam M. Satisfaction with care in post-stroke patients undergoing a telerehabilitation programme at home. Journal of telemedicine and telecare 2008;14(5):60.
23. Redzuan N, Engkasan J, Mazlan M, Abdullah S. Effectiveness of a video-based therapy program at home after acute stroke: a randomized controlled trial. Archives of physical medicine and rehabilitation 2012;93(12):83.
24. Wolf SL, Sahu K, Bay RC, Buchanan S, Reiss A, Linder S. The HAAPI (Home Arm Assistance Progression Initiative) trial: a novel robotics delivery approach in stroke rehabilitation. Neurorehabilitation and neural repair 2015;29(10):68.
25. Cimarolli VR, Reinhardt JP, Horowitz A. Perceived overprotection: support gone bad. J Gerontol B Psychol Sci Soc Sci 2006 Jan;61(1):S18-S23. [Medline: 16399945]
26. Visser-Meily A, Post M, van de Port I, Maas C, Forstberg-Wärleby G, Lindeman E. Psychosocial functioning of spouses of patients with stroke from initial inpatient rehabilitation to 3 years poststroke: course and relations with coping strategies. Stroke 2009 Apr;40(4):1399-1404. [doi: 10.1161/STROKEAHA.108.516682] [Medline: 19095973]
27. Hjelm N. Benefits and drawbacks of telemedicine. Journal of telemedicine and telecare 2005;11(2):70.
28. Akbik F, Hirsch J, Chandra RV, Frei D, Patel A, Rabinov J. Telestroke-the promise and the challenge. Part two-expansion and horizons. Journal of neurointerventional surgery 2017;9(4):5.
29. Sarfo F, Adamu S, Awuah D, Sarfo-Kantanka O, Ovbiagele B. Potential role of tele-rehabilitation to address barriers to implementation of physical therapy among West African stroke survivors: A cross-sectional survey. Journal of the neurological sciences 2017:8.
30. Donoso BEV, McCoy S, Fechko A, Price R, Gilbertson T, Moritz C. Preliminary investigation of an electromyography-controlled video game as a homeprogram for persons in the chronic phase of stroke recovery. Archives of physical medicine and rehabilitation 2014;95(8):9.
31. Fauchier L, Sadoul N, Kouakam C, Briand F, Chauvin M, Babuty D, et al. Potential cost savings by telemedicine-assisted long-term care of implantable cardioverter defibrillator recipients. Pacing Clin Electrophysiol 2005 Jan;28 Suppl 1:S255-S259. [doi: 10.1111/j.1540-8159.2005.00071.x] [Medline: 15683509]
32. Thaker DA, Monypenny R, Olver I, Sabesan S. Cost savings from a telemedicine model of care in northern Queensland, Australia. Med J Aust 2013 Sep 16;199(6):414-417. [Medline: 24033216]
33. Zhai Y, Zhu W, Cai Y, Sun D, Zhao J. Clinical- and cost-effectiveness of telemedicine in type 2 diabetes mellitus: a systematic review and meta-analysis. Medicine (Baltimore) 2014 Dec;93(28):e312 [FREE Full text] [doi: 10.1097/MD.0000000000000312] [Medline: 25526482]
34. Whitten PS, Mair FS, Haycox A, May CR, Williams TL, Hellmich S. Systematic review of cost effectiveness studies of telemedicine interventions. BMJ 2002 Jun 15;324(7351):1434-1437 [FREE Full text] [Medline: 12065269]
35. Laver K, Schoene D, Crotty M, George S, Lannin N, Sherrington C. Telerehabilitation services for stroke. Cochrane Database Syst Rev 2013 Dec 16(12):CD010255. [doi: 10.1002/14651858.CD010255.pub2] [Medline: 24338496]
36. Egger M, Davey SG, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634 [FREE Full text] [Medline: 9310563]

Abbreviations

ARAT: Action Research Arm Test
FM-UE: Fugl-Meyer Upper Extremity
HRQoL: health-related quality of life

SF: short-form
SMD: standardized mean difference
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 24.04.18; peer-reviewed by A Ibrahim Abushouk, M Dama, C Jacob, T Poder; comments to author 10.05.18; revised version received 18.07.18; accepted 22.07.18; published 26.10.18.

Please cite as:
Tchero H, Tabue Teguo M, Lannuzel A, Rusch E
Telerehabilitation for Stroke Survivors: Systematic Review and Meta-Analysis
J Med Internet Res 2018;20(10):e10867
URL: http://www.jmir.org/2018/10/e10867/
doi:10.2196/10867
PMID:30368437

Original Paper

The Use of Web-Based Technologies in Health Research Participation: Qualitative Study of Consumer and Researcher Experiences

Patrick Cheong-Iao Pang[1], MSc, PhD, CHIA; Shanton Chang[1], PhD; Karin Verspoor[1,2], PhD; Ornella Clavisi[3,4], MPH

[1]School of Computing and Information Systems, The University of Melbourne, Parkville, Australia
[2]Health and Biomedical Informatics Centre, The University of Melbourne, Parkville, Australia
[3]Musculoskeletal Australia, Elsternwick, Australia
[4]Australian and New Zealand Musculoskeletal Trials Network, Melbourne, Australia

Corresponding Author:
Patrick Cheong-Iao Pang, MSc, PhD, CHIA
School of Computing and Information Systems
The University of Melbourne
Doug McDonell Building (Building 168)
The University of Melbourne
Parkville, 3010
Australia
Phone: 61 3 9035 6066
Email: mail@patrickpang.net

Abstract

Background: Health consumers are often targeted for their involvement in health research including randomized controlled trials, focus groups, interviews, and surveys. However, as reported by many studies, recruitment and engagement of consumers in academic research remains challenging. In addition, there is scarce literature describing what consumers look for and want to achieve by participating in research.

Objective: Understanding and responding to the needs of consumers is crucial to the success of health research projects. In this study, we aim to understand consumers' needs and investigate the opportunities for addressing these needs with Web-based technologies, particularly in the use of Web-based research registers and social networking sites (SNSs).

Methods: We undertook a qualitative approach, interviewing both consumer and medical researchers in this study. With the help from an Australian-based organization supporting people with musculoskeletal conditions, we successfully interviewed 23 consumers and 10 researchers. All interviews were transcribed and analyzed with thematic analysis methodology. Data collection was stopped after the data themes reached saturation.

Results: We found that consumers perceive research as a learning opportunity and, therefore, expect high research transparency and regular updates. They also consider the sources of the information about research projects, the trust between consumers and researchers, and the mobility of consumers before participating in any research. Researchers need to be aware of such needs when designing a campaign for recruitment for their studies. On the other hand, researchers have attempted to establish a rapport with consumer participants, design research for consumers' needs, and use technologies to reach out to consumers. A systematic approach to integrating a variety of technologies is needed.

Conclusions: On the basis of the feedback from both consumers and researchers, we propose 3 future directions to use Web-based technologies for addressing consumers' needs and engaging with consumers in health research: (1) researchers can make use of consumer registers and Web-based research portals, (2) SNSs and new media should be frequently used as an aid, and (3) new technologies should be adopted to remotely collect data and reduce administrative work for obtaining consumers' consent.

(J Med Internet Res 2018;20(10):e12094) doi:10.2196/12094

KEYWORDS

research subjects; consumer behavior; research design; social networking; patient portals; registries

Introduction

Background

Health consumers, who are users or potential users of health care services (eg, patients, families, carers, and other support people) [1], play a valuable role in health research [2,3]. They are not simply research participants these days; they are involved in research in various ways, including providing input into research design, joining advisory committees, raising public awareness, and disseminating research information [4,5]. Recruitment and engagement of consumers have been ongoing critical but challenging tasks for researchers [6-10]. A patient-centered approach has recently been advocated so that consumers can make informed decisions about whether they should participate in research [11]. However, it is often not easy to identify suitable tools to reach out to health consumers and to convey the research goals.

Recently, researchers have turned to Web-based technologies such as Web-based patient registers and social networking sites (SNSs) to recruit research participants [9,12]. Research suggests that these technologies have numerous advantages for conducting research studies, such as maintaining a list of enthusiastic potential participants [13-15], reducing recruitment cost [12,16], and providing the ability to identify hard-to-reach participants [14,17]. However, there is limited work investigating the engagement of research participants. In fact, it has been reported that there is a need to discover more clear and effective approaches for improving participant’s engagement in health research [18,19]. In addition, it is unclear whether or specifically how Web-based technologies can help with recruitment, participation, and engagement. As such, more research is needed to identify the appropriate use of Web-based technologies for these purposes.

Our review of the literature has shown that understanding and meeting the needs of health consumers can improve their experience and engagement in digital health applications [20-22] and thereby result in positive outcomes [23-25]. It is important to investigate why and how the consumers make decisions and subsequently respond to their requirements [26]. We have also identified that researchers may not fully leverage always-on Web environment and human online interactions to engage end users [27]. Given this background, we aim to investigate consumers’ needs for participating in health research and to identify opportunities to engage participants using relevant technologies.

In this study, we used a qualitative approach and obtained insights from both consumers and researchers in the area of musculoskeletal (MSK) research. We chose MSK research as our focus because many projects in this research area heavily involve consumer participants for improving their quality of life [28-31]. Therefore, this cohort can give us rich feedback about their motivations and experience about participating in research. Although our data were collected from a limited group, we expect that the insights obtained will not only be applicable to the MSK research community but also more generally to health researchers who require significant consumer involvement.

We carried out this research with the help of an Australian MSK-supporting nonprofit organization because such organizations are often considered to be an effective contact point with health consumers [32]. Through their connections, we interviewed consumers with MSK conditions for their thoughts and feedback about participating in health research as well as researchers who have conducted MSK research to understand their current practices for engaging with participants. On the basis of the interview data, we identified gaps in the current practice and proposed new directions for using technologies to improve consumer participation and engagement. We expect our study to bring a different perspective to the adoption of Web-based technologies in conducting health research, with the understanding of consumers and their needs in mind. In addition, we hope to open up a discussion about applying similar strategies to health research in other areas.

Aims

The following 3 main research questions (RQ) comprise the aims of this study:

> RQ1: What do consumers need and expect to gain from participating in health research projects?
>
> RQ2: What current practice is used by researchers in engaging with consumers for their research projects, and what are the challenges?
>
> RQ3: How can Web-based technologies be employed to improve engagement of research participants?

The study involves both health consumers and researchers to answer these questions. For consumers, we seek to systematically investigate their motivations for and decision-making processes about taking part in research to guide researchers to consider participants’ needs in their research design and to enhance the experience of people getting involved in academic research. On the other hand, we suggest that this study provides an opportunity for researchers to reflect on their research processes and the difficulties in engaging consumers in their studies. By comparing the needs of consumers and the current practices of researchers, we identify the strategies and opportunities to align research with participants’ expectations through the adoption of appropriate technologies.

Methods

Methodology

This study adopts a qualitative approach to collect and analyze data. Semistructured interviews were conducted with both consumers and researchers. In-person interviews were preferred, but phone interviews were also arranged for people who could not come on-site because of time and mobility issues. All interview sessions were conducted in the presence of authors PP and OC. This research was approved by the human research ethics committee of the University of Melbourne (Approval ID: 1648346.1).

All interviews were recorded and transcribed for data analysis. Transcripts were processed using the thematic analysis approach [33,34]. The data analysis involved reading the transcripts, mapping key ideas into codes with the open coding process, and summarizing codes into themes iteratively [35]. In the process, the number of participants connected to each theme was tracked for verifying the generality of themes. After generalizing themes, the list of themes was verified by another person in our department to assure consistency, accuracy, and quality [36].

Recruitment

For this study, consumer participants were recruited through Musculoskeletal Australia, which is an Australian nongovernmental organization that supports people with MSK conditions nationwide. For the recruitment of other researchers in this study, we sent out invitations to their contact lists and the university's mailing list and then reached out to the ones who responded and had previously conducted research in the area of MSK conditions. The recruitment of both consumers and researchers continued until we reached data saturation [37,38], that is, there were no new themes emerging from the collected data. Interviewees did not receive any incentives for participating in our study.

Interview Design

For the consumer interviews, we structured our questions around their motivations, information needs, and experience of participating in research projects. In addition, the interviews focused on their views on researchers and expectations of getting involved in future research. Table 1 outlines the interview questions used in our study.

In the interviews of researchers, we started with a number of general questions to understand what research directions they were pursuing. Then, we asked about different aspects of their research projects, ranging from recruitment, research methods, and difficulties, to costs and incentives. In addition, we investigated the channels and the types of media used for promoting their research. Table 2 shows the interview questions for researchers.

Participants

Representing consumers, we completed interviews with 23 people having MSK conditions. The mean age of the participants was 51 years (SD 15.7 years, range 15-72 years). About half of our participants (48%, 11/23) reported that they had previously taken part in academic research. Table 3 summarizes the demographics of our participants.

Table 1. The list of interview questions for consumers.

Category	Questions
Demographic	• Age, gender, remoteness, work status, and conditions
Motivation	• What makes you want to find out more about academic research? • Why is a research project of interest to you? • What do you expect to gain from participating in a research project?
Opportunities to get involved in research	• Where do you get information about particular research projects? • How do you find research that is relevant to you? • What websites or tools do you use to find research? • What are your search criteria?
Experience	• What is your overall experience with getting involved in a research study? • What are the difficulties and challenges? • How do you think technologies can improve your participation?

Table 2. The list of interview questions for researchers.

Category	Questions
General information	• What kinds of studies do you normally carry out? • Will the participants need to be involved for a long term? • What is the typical time commitment required from them?
Recruitment	• What are the difficulties you face when recruiting participants for your research? • What are the factors that lead to a successful recruitment? • What information do you provide to recruit potential participants? • How do you screen relevant or appropriate participants? • How much time do you spend on recruitment? • What incentives do you give to your participants? • How much cost do you spend on marketing?
Channels	• What channel do you currently use to recruit? • How do you think that technologies can help to improve the participant recruitment and engagement in health research? • Would you able to use materials such as images, photos, or videos to recruit participants? Why or why not? • Have you considered using a consumer register or a potential participant database for recruitment?

Table 3. Demographics of consumer participants in our study (n=23).

Demographics	n (%)
Gender	
Male	5 (22)
Female	18 (78)
Living area	
Metro	14 (61)
Rural	9 (39)
Work status	
Full-time	6 (26)
Part-time	5 (22)
Unemployed	12 (52)

Representing researchers, we interviewed 10 researchers who study various aspects related to MSK conditions. These included clinical research in back pain and spinal pain, the quality of life with osteoarthritis, observational studies after surgery, as well as epidemiology research. All researchers were based in Australia and possessed a minimum academic level B (Lecturer; equivalent to Assistant Professor in the US system) position. They played the role of chief investigator of academic research projects and directly led ongoing or past studies in the MSK research area.

Results

Overview

This section presents the data collected from the interviews with both consumers and researchers. We further organize the results into 2 subsections: consumer needs and researcher strategies.

Consumer Needs

This section reports on the consumers' needs (*consumer needs, CN*), which can be further broken down into 4 subthemes: research as learning opportunities (CN1), research transparency and updates (CN2), trustworthiness (CN3), and mobility (CN4). The corresponding consumer identifier is listed after each quote.

CN1: Research as Learning

Many of our consumer participants, particularly those with less access to information sources, viewed researchers as being on the frontier of science, and, therefore, they hoped to learn something from research. Participants reported that they were enthusiastic about taking part in research because they would have a chance to enquire about the latest remedies and treatments that might be useful to them. In addition, research was a learning opportunity for them to obtain new knowledge about their conditions:

> *For me it is the possibility of new information on effective treatments becoming available to me and I do a lot of my own looking around but—or just trying to keep on top of new developments I guess.* [C4]

> *Research is one of the things that people want to know about because it gives them a sense as a set of hope[s] that things might be different for them in the future.* [C5]

CN2: Research Transparency and Updates

Consumers indicated that receiving more clear information about the purpose, the scope, and the protocols of a research project is essential for them to get involved. As people with MSK conditions sometimes found participating in research activities difficult in terms of access or mentally demanding, they tended to choose carefully beforehand and only invested in research that was relevant and beneficial to them:

> *[Researchers need to] be explicit about what they expect from participants. Because, as I said, there's a lot of anxiety with a lot of people with chronic illness about how much they have to give to anything energy wise...* [C5]

On the other hand, consumers wanted to receive feedback regarding the progress of the study as well as the final results. Furthermore, they were disappointed when they did not receive any communication regarding research that they had participated in. As noted by our interviewees, this appears to be a very common issue:

> *In the end, I didn't ever get to hear any results back. I've got obviously, thanked for my participation, but it was disappointing that I never heard anything back.* [C21]

CN3: Information Sources and Trustworthiness

A large portion of our consumers suggested that they would not proactively search for relevant research to participate in. Instead, it is essential for someone to inform them about eligible studies. Many of our participants highlighted the importance of health professionals such as general practitioners, rheumatologists, and physiotherapists for introducing or actively referring them to research projects. One participant said, "I probably don't go about looking for something until I hear about something" (C1).

In addition, the participants reported that the trustworthiness of information sources was critical for them to consider regarding whether the research is worthwhile to participate in. Some people emphasized that they would not take part in any pharmaceutical or marketing research. They also had concerns about how to find genuine research on the internet, for example, from Google. In this case, community services, support groups, and health professionals played an important role to provide information to potential research participants:

> *I'd probably go by word of mouth and ask other people or other clinicians or something like that rather than just go randomly onto Google.* [C3]
>
> *I'd rather someone I trust to tell me where to go [for research]...* [C7]

CN4: Mobility and Rural Locations

In the consultations with our participants, we understood that people with MSK conditions needed extra considerations for their conditions and special needs in the research design. For instance, some older participants suggested that they preferred physical contact and building a social connection with the researchers. They tended to talk more and to have more interactions in the progress. These factors need to be considered before adopting the use of technologies:

> *I think for some of the older-older people prefer to have a face to face conversation. They like to talk to people, they like that option of being able to have a chat.* [C4]

Besides, a significant ratio of the consumers had low mobility, which limited their ability to travel for academic studies. This was a particular problem for people living in rural areas and made them feel isolated. As a result, technology aids were suggested to address these limitations and to make research more accessible. One participant expressed, "I've got to weigh up the time and the effort involved physically and how much pain it would cost me... the online stuff is fine" (C8).

Researcher Strategies

This section presents the strategies used by researchers (*researcher strategies, RS*) to assist with recruitment and to maintain consumers' engagement for their research. Overall, 4 main strategies have been identified: establishing rapport with consumers (RS1), handling changes in research (RS2), using technologies (RS3), and designing research for the participants (RS4). The corresponding researcher identifier is shown after each selective quote.

RS1: Establishing Rapport With Consumers

Researchers pointed out the importance of establishing a connection with consumers at the beginning of the recruitment to engage them with the research project. Although many MSK research projects recruited through clinicians and health professionals, their interaction with the potential participants was crucial:

> *If that person [the clinician] who has sort of clinical responsibility for the patient is into it and is sort of an advocate for the research, then I think that's what—in some ways the recruitment lives or dies by that. That person is really sort of going into bat for your study...* [R8]

It was also helpful to involve consumers in the very early stages of the research, even in the design of the trial, so that they could be informed about the purpose of the research and the benefits to them. It was worthwhile for consumers to learn more about the research and the rationale behind it:

> *[The] number one is just engaging the consumer right at the beginning, so that we understand that we're answering and addressing the question that's of real importance to consumers, even in the design of the trial...trying to work out how to explain to consumers why it's so important and it is a good investment of their time.* [R5]

Some researchers found that the time spent by research team members to talk and build up a relationship with the potential participants was a good investment in some studies:

> *That person [the staff member] will take the time to have a chat with them about the project, and I think when you establish the rapport like that...I think if you had that connection they are more likely to consent to being part of the study.* [R2]

Conversely, poor communication would endanger the recruitment. This also highlights the needs of researchers and facilitators to manage the relationship with consumers:

> *We have had circumstances where the communication between the research assistant and the participant wasn't great, and so then participants, they get disappointed and they decided to drop the study...The studies that we've been most successful in keeping patients in the study, are studies where we actually have someone actually spending time talking to patients and that particular person is someone who develops a relationship with the participants.* [R1]

With these challenges in recruiting participants by the researchers themselves, they used patient groups or consumer organizations to help with the recruitment. These organizations had direct connections to the consumers. As a researcher suggested, "Going through a consumer organization or a professional body is probably a better way to go" (R2).

RS2: Handling Changes in Research

Changes are inevitable in research. Researchers found that it was a challenge to go back to the participants to request more information or to obtain consent when there were changes in their research:

> *If you just list one very specific question, then if you're going to re-analyze that data with different questions, or if you need to go back to the twins [the participants of that study] and re-interview them, then you need to go back [sic] and ask for permission again.* [R1]

Some researchers reported on their experience of obtaining multilevel consent, which allowed consumers to choose the level of involvement such that they were comfortable throughout the research. In this way, the experience of the participants was enhanced, and the administrative burdens of obtaining additional consents were lowered:

> *I suppose we're talking about really, the consent process here. So in a way, what would be best practice*

> *would be multi-level consent. So, you could have the tick all box, which is, "I'm happy to be contacted by anybody about anything at any point," and everything backwards from there, up to, "Don't contact me about anything, I'll contact you," and everything in between. And, in a way, I suppose offering that spectrum is the most appropriate thing to do and, in some ways, supporting people to make an informed decision about where they want to sit on that spectrum.* [R6]

RS3: The Use of Technologies

Researchers agreed that media such as images and videos were more powerful. As such, they started to use SNSs such as Facebook and YouTube to help promote their research. Text information can carry a large amount of precise information, but often it is not sufficiently engaging. In contrast, short videos could help to deliver the basic message and the background of the research. However, researchers also mentioned that they did not know where to obtain appropriate images or how to modify the images for the best results on SNSs. For videos, there were also reports on having difficulties with editing videos and producing satisfactory output quality:

> *We're very much happy with technology that still involves images, for example, so Skype and tele-rehabilitation is a good example, patients and participants tend to be quite happy with that. Technologies that don't involve images, or such as just text and email, and that sort of thing, I think, they tend to be a problem.* [R1]
>
> *One of my concerns with it that there is so little (video and audio)—it's all written information. Like there's no other way of receiving that information. And video and audio and all that sort of stuff is so useful.* [R3]
>
> *Well, why don't we start a three-minute video placed on YouTube, explaining what it is...And, in under 12 months, we've had about 26,000 views of that video.* [R6]
>
> *The first of these things are videos, like, YouTube videos, to express things you are researching, using a language that is more easy to understand and provide a bit of background about the condition or the things you are working on.* [R2]

RS4: Designing Research for the Participants

Time commitment was one of the issues identified by MSK researchers in their research. The time of the participants was precious, and the researchers tried to minimize the time required to travel for data collection. Some researchers used Web-based recruitment tools and survey platforms. However, for electronic data collection, it was important that surveys be kept to a reasonable length to avoid participants dropping out in the middle of a lengthy session. Such details in the research design directly affected the dropout rate of participants:

> *So, in terms of collecting the data, we try to keep this data collection session as quick as we can, anything over one hour becomes a problem...if it's a survey, an electronic survey, it needs to be much less than that, it needs to be no longer than 20 minutes, otherwise it becomes a major issue, the participants tend to drop out of the study.* [R1]

In addition to time commitment, travel distance was reported as another issue that affected participation. People living in remote areas had more difficulties gaining access to studies. Researchers described the use of Web-based tools to help these people to participate in research and to collect data:

> *We had a lot of rural and even remote people that we had. We did virtual focus groups, so an online focus group, and we had the face to face ones with people living in metro.* [R2]

Another challenge pointed out by our interviewees was that arranging meetings and focus groups for participants was difficult. Participants had different schedules and might not have a common time to meet. This was particularly true for focus group research because it required a minimum number of people to be present:

> *In order to hold a focus group, you have to have between six to eight minimum participants who can come on the same day at the same location at the same time. So, it was a process of asking people if they could give me [sic] of their availability and then trying to match a general availability with a specific case didn't always work out. Yeah, so getting participants with similar availability was the main complication for focus group research.* [R7]

Discussion

Principal Findings

Through the interviews, we have gained insight into the motivations and the needs of consumers who participate in research projects. We have also learned that researchers have already identified some of the key issues and are adopting strategies to keep consumers engaged. For example, the following points were noted:

- Researchers establish a good rapport with participants, which further fosters more learning opportunities for research participants (RS1→CN1).
- Researchers invest the effort to build an ongoing relationship with consumers from the beginning of the recruitment, which helps to create trust and confidence in consumers' hearts (RS1→CN3).
- Researchers visit and request new consent again when there are changes in their research, which allows greater transparency and updates relevant to participants (RS2→CN2).
- Researchers have taken up new technologies to help consumers who live in rural locations or have low mobility participate in the research (RS3→CN4).
- Researchers have started to design their research to be better tailored for the potential participants, such as using online focus groups, splitting meetings, and reducing the time required for taking a session (RS4→CN4).

Figure 1. Key consumer needs and researcher strategies identified through the research.

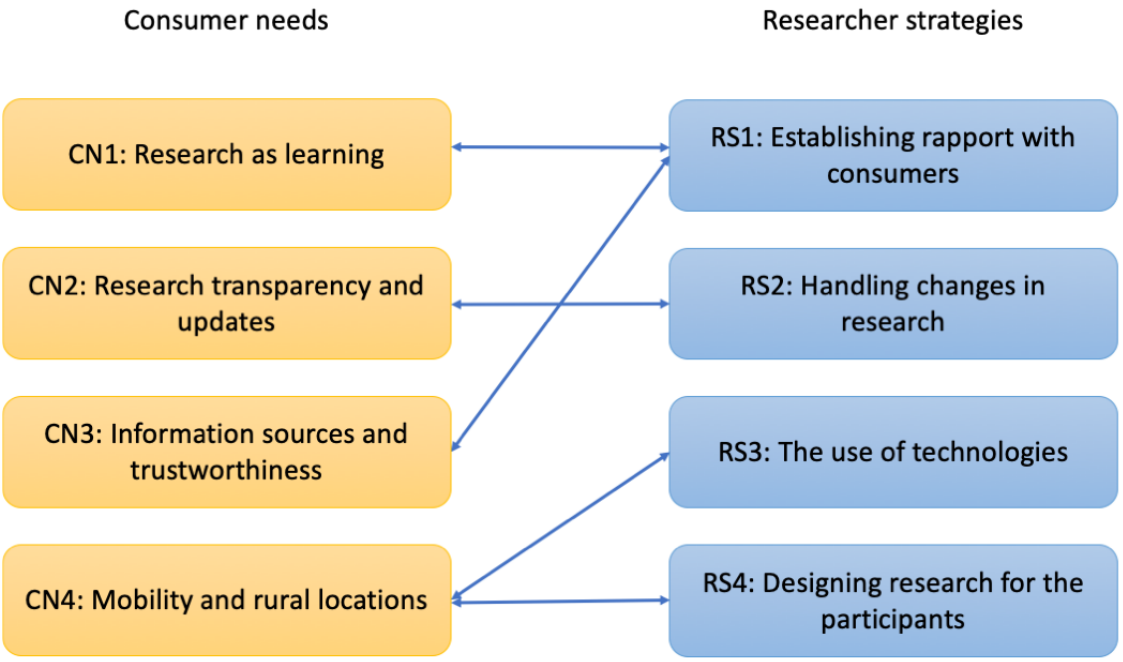


Figure 2. Opportunities for supporting consumer needs with Web-based technologies.

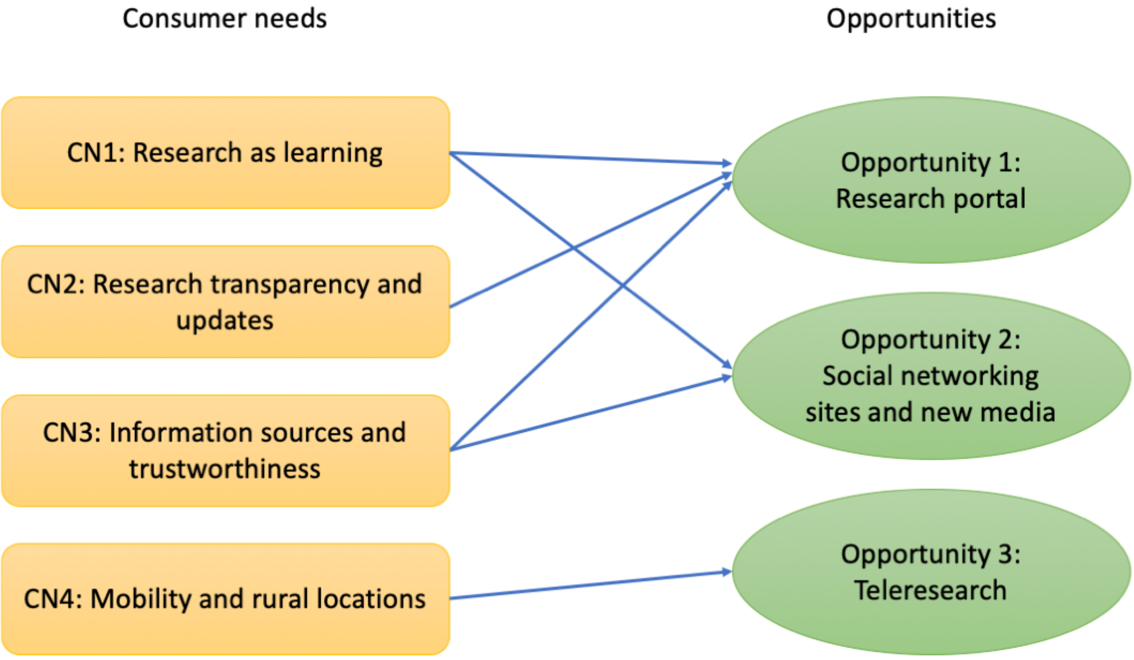


Figure 1 summarizes consumer needs and how these needs are being addressed by current researcher strategies. Although effort has been devoted to engage consumers in research activities, we identify 3 areas that can address their needs better through the use of Web-based technologies (Figure 2). These will be elaborated in the later parts of this section.

Opportunity 1: Research Portal

Past research has shown that patient registers can effectively recruit participants [9,10,13,39,40]. Taking a step further, we propose that such registers can be augmented with streamlined research portals that further address different consumer needs. A consumer-facing research portal can provide multiple functions, including promoting research opportunities, allowing electronic participant recruitment, and sending notifications and updates to relevant users through the Web.

As suggested by our results, consumers tend to have passive information-seeking behavior [41,42] for acquiring information about health research. In other words, people do not realize the needs of getting information and instead rely on other people to trigger them to start seeking information. As a result, they do not notice that research opportunities actually exist unless they get informed. In this case, a research portal can store contact information of consumers and enable researchers to notify them of new studies. This will ignite their interest to learn more about the new research and consider participation (CN3).

In addition, a research portal can show a list of research projects and present the current status, updates, and even recent publishable results of such projects. This approach can improve research transparency and provide timely updates to consumers (CN2). This can fulfill their needs for getting news and updates at their convenience and eventually help them to learn more about the study. A research portal can be seen as another credible resource for consumers, which allows them to understand and acquire the latest knowledge about their conditions and thereby addresses the learning needs (CN1).

On the other hand, consumer organizations can help with engaging participants in health research [43]. First, consumer groups can help to advocate the use of patient registers and research portals as well as to disseminate the research information. Also told by researchers (RS1), working with consumer groups is helpful for their research, which results in establishing rapport and trust with consumers (CN3). Besides, the information on consumer-faced health portals often requires a higher level of literacy [44,45] as there are disparities of the literacy and the knowledge levels between authors (such as researchers and clinicians) and consumers. We anticipate that research portals might have similar issues. In this case, consumer organizations can play an intermediate role to review research material and interpret the content in an easy-to-understand manner for consumers. This will facilitate the learning, understanding, and involvement in health research (CN1).

Furthermore, a research portal can act as a credible information source, which helps to mitigate the issues of confidence and trustworthiness (CN3). A Web-based research portal can make the identity of the project owners and the administrators more transparent so that visitors can know of the nature and the ownership of the portal (eg, *is it noncommercial?* or *is it supported by a pharmaceutical company?*). The branding of the portals can also help to build up the confidence and eventually improve consumer satisfaction [46]. Finally, a centralized-managed research portal can be easier managed and secured by information and technology professionals who have the expertise in the operational and cybersecurity perspectives that health researchers do not normally have.

Opportunity 2: Social Networking Sites and New Media

As reported by researchers (RS3), SNSs are being used to carry out recruitment, which is consistent with other literature [47-49]. Despite the wide use of SNSs, it remains a challenge to recruit research participants properly through these sources. It is reported that extra consideration of recruitment design as well as technical work are needed to prevent repeated attempts from the same person and biased samples when research is advertised on these social networking platforms [48,49]. Additionally, SNS users are not evenly distributed across different age levels and ethnicity background [50]. Therefore, we suggest that the use of SNSs should be only a part of the whole recruitment, unless the targeted cohort is specifically people who use SNSs heavily.

The recent trend of using multimedia content on SNSs brings new opportunities to researchers. Researchers understand that using images and video clips has advantages over the classical text-based material. Recent papers have pointed out that posts with rich content (eg, images and videos) capture more attention on SNSs [51-53], which researchers can leverage to promote their research more effectively. On the other hand, videos often provide a shorter and clearer presentation of information and can employ visual aids to help watchers to learn and understand the research context (CN1). The use of human faces in videos helps to build a relationship with the audience (RS1). This gives a starting point to establish the trust (CN3) and the rapport (RS1) with the potential participants.

However, the use of new media on SNSs creates new technical burdens for researchers. As highlighted in the results (RS3), researchers found it difficult to use or manipulate multimedia on SNSs. This aligns with previous literature that using SNSs for health research is resource-demanding [12,54]. To overcome the issue, we suggest a *social media kit* can be designed for researchers to provide sources of properly licensed (eg, licensed under Creative Commons that allows the reuse of materials) images and videos that are technically usable on SNSs. In addition, more instructions and tutorials can be offered for effectively editing and using new media on SNSs. Universities and institutions can consider setting up a team of social media specialists for helping researchers to broadcast their research on the internet with low overheads and matching the audience with a propensity for relevant research topics.

Opportunity 3: Teleresearch

With the growing momentum of faster internet networks and mobile and wearable technologies, researchers can consider using the latest development of teleresearch technologies to collect data and conduct research activities. For example, mobile apps can be used to log user activities and could be an alternative to conducting diary studies; videoconferencing software (eg, Skype and Zoom) can be used as a channel for group interviews and focus groups. Recent work about wearable gadgets such as smart socks [55] allows researchers to collect body measurements remotely. These technologies provide participant-friendly solutions for diverse needs of time, place, and mobility (CN4).

Additionally, technologies can assist with the change in the research process (RS2). The latest research suggests that using dynamic consent to collect consent digitally can help consumers to make decisions for their participation in research and reduce the administrative burden of researchers [56,57]. For instance, dynamic consent can split the entire consent of a study into a few consent items, and users will only be prompted to give consent on the fly when relevant experiments are performed. As such, the system can notify the participants to provide additional permissions online without the need to contact them by physical means. This simplifies the complexity on the researcher side handling changes in the research. In addition, participants can choose to transfer their digital consent to third parties, which enables easier collaborations and data reuse across research teams. Recent trials have started to explore the deployment of digital consent and its efficacy in various settings [58].

Limitations

We acknowledge several limitations of our study. Both the samples of consumers and researchers may not represent the entire population because of the relatively small number of subjects. For consumers, the sample was biased in favor of women and older cohorts. However, this can be explained by the fact that more women and older people suffer from MSK problems [59]. In addition, the recruitment of this research was conducted through a single consumer organization, which might affect the diversity of the sample. Finally, the adoption of technologies may impact the participant cohort with lower digital and health literacy. More research in the future should investigate the impact on this cohort.

Conclusions

On the basis of the interviews with consumers and researchers, we summarize 4 major types of consumer needs as well as 4 strategies used by researchers for engaging participants. On the basis of these findings, we argue that 3 areas of Web-based technologies can be employed to assist in addressing consumer needs and engaging with research participants: research portals, SNSs, and teleresearch. Additionally, our research outcomes lead to a better understanding of human participants and offer an opportunity to reflect on the research design. The analysis presented in this paper is not just relevant to a single discipline but is also applicable to other types of health research.

Acknowledgments

The authors acknowledge the support from Musculoskeletal Australia for this study. They are grateful to Sara Servodio and Sarah-Jane Dyrenfurth for contacting the participants and organizing interviews. In addition, they thank Juanru Liu for assisting with the analysis of the interview transcripts. Finally, they are appreciative of the consumers and the researchers who contributed their time and insights to this study.

Conflicts of Interest

None declared.

References

1. National Health and Medical Research Council, Consumers Health Forum of Australia. Statement on consumer and community involvement in health and medical research. 2016. URL: https://nhmrc.gov.au/about-us/publications/statement-consumer-and-community-involvement-health-and-medical-research#block-views-block-file-attachments-content-block-1 [WebCite Cache ID 72lRSiol2]
2. Polgar S, Thomas S. Introduction to Research in the Health Sciences. Philadelphia, PA: Churchill Livingstone, Elsevier; 2007.
3. Synnot AJ, Cherry CL, Summers MP, Stuckey R, Milne CA, Lowe DB, et al. Consumer engagement critical to success in an Australian research project: reflections from those involved. Aust J Prim Health 2018 Dec;24(3):197-203. [doi: 10.1071/PY17107] [Medline: 29875031]
4. Shippee ND, Domecq GJ, Prutsky LG, Wang Z, Elraiyah TA, Nabhan M, et al. Patient and service user engagement in research: a systematic review and synthesized framework. Health Expect 2015 Oct;18(5):1151-1166 [FREE Full text] [doi: 10.1111/hex.12090] [Medline: 23731468]
5. INVOLVE. 2018. What is public involvement in research? t URL: http://www.invo.org.uk/find-out-more/what-is-public-involvement-in-research-2/ [WebCite Cache ID 72l50kez4]
6. Bower P, Wallace P, Ward E, Graffy J, Miller J, Delaney B, et al. Improving recruitment to health research in primary care. Fam Pract 2009 Oct;26(5):391-397. [doi: 10.1093/fampra/cmp037] [Medline: 19549623]
7. Fenner Y, Garland SM, Moore EE, Jayasinghe Y, Fletcher A, Tabrizi SN, et al. Web-based recruiting for health research using a social networking site: an exploratory study. J Med Internet Res 2012;14(1):e20 [FREE Full text] [doi: 10.2196/jmir.1978] [Medline: 22297093]
8. Patel MX, Doku V, Tennakoon L. Challenges in recruitment of research participants. Adv Psychiatr Treat 2018 Jan 02;9(3):229-238. [doi: 10.1192/apt.9.3.229]
9. Sully BG, Julious SA, Nicholl J. A reinvestigation of recruitment to randomised, controlled, multicenter trials: a review of trials funded by two UK funding agencies. Trials 2013;14:166 [FREE Full text] [doi: 10.1186/1745-6215-14-166] [Medline: 23758961]
10. Fazzino TL, Rose GL, Pollack SM, Helzer JE. Recruiting US and Canadian college students via social media for participation in a web-based brief intervention study. J Stud Alcohol Drugs 2015 Jan;76(1):127-132. [doi: 10.15288/jsad.2015.76.127]
11. Sacristán J, Aguarón A, Avendaño-Solá C, Garrido P, Carrión J, Gutiérrez A, et al. Patient involvement in clinical research: why, when, and how. Patient Prefer Adherence 2016 Apr 27;10:631-640. [doi: 10.2147/PPA.S104259]
12. Whitaker C, Stevelink S, Fear N. The use of Facebook in recruiting participants for health research purposes: a systematic review. J Med Internet Res 2017 Aug 28;19(8):e290 [FREE Full text] [doi: 10.2196/jmir.7071] [Medline: 28851679]

13. Lovato L, Hill K, Hertert S, Hunninghake DB, Probstfield JL. Recruitment for controlled clinical trials: literature summary and annotated bibliography. Control Clin Trials 1997 Aug;18(4):328-352. [doi: 10.1016/S0197-2456(96)00236-X] [Medline: 9257072]
14. McKinstry B, Sullivan FM, Vasishta S, Armstrong R, Hanley J, Haughney J, et al. Cohort profile: the Scottish Research register SHARE. A register of people interested in research participation linked to NHS data sets. Br Med J Open 2017 Dec 01;7(2):e013351 [FREE Full text] [doi: 10.1136/bmjopen-2016-013351] [Medline: 28148535]
15. Larsen ME, Curry L, Mastellos N, Robb C, Car J, Middleton LT. Development of the CHARIOT Research Register for the Prevention of Alzheimer's Dementia and Other Late Onset Neurodegenerative Diseases. PLoS One 2015;10(11):e0141806 [FREE Full text] [doi: 10.1371/journal.pone.0141806] [Medline: 26599631]
16. Heffner JL, Wyszynski CM, Comstock B, Mercer LD, Bricker J. Overcoming recruitment challenges of web-based interventions for tobacco use: the case of web-based acceptance and commitment therapy for smoking cessation. Addict Behav 2013 Oct;38(10):2473-2476 [FREE Full text] [doi: 10.1016/j.addbeh.2013.05.004] [Medline: 23770645]
17. King DB, O’Rourke N, DeLongis A. Social media recruitment and online data collection: a beginner’s guide and best practices for accessing low-prevalence and hard-to-reach populations. Can Psychol 2014;55(4):240-249. [doi: 10.1037/a0038087]
18. Domecq JP, Prutsky G, Elraiyah T, Wang Z, Nabhan M, Shippee N, et al. Patient engagement in research: a systematic review. BMC Health Serv Res 2014 Feb 26;14:89 [FREE Full text] [doi: 10.1186/1472-6963-14-89] [Medline: 24568690]
19. Forsythe LP, Szydlowski V, Murad MH, Ip S, Wang Z, Elraiyah TA, et al. A systematic review of approaches for engaging patients for research on rare diseases. J Gen Intern Med 2014 Aug;29 Suppl 3:S788-S800 [FREE Full text] [doi: 10.1007/s11606-014-2895-9] [Medline: 25047393]
20. Pang PC, Verspoor K, Pearce J, Chang S. Better Health Explorer: Designing for Health Information Seekers. In: Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction (OzCHI '15).: ACM Press; 2015 Presented at: OzCHI'15; December 7, 2015; Melbourne, Australia p. 588-597. [doi: 10.1145/2838739.2838772]
21. Lee K, Hoti K, Hughes JD, Emmerton LM. Consumer Use of “Dr Google”: a survey on health information-seeking behaviors and navigational needs. J Med Internet Res 2015;17(12):e288 [FREE Full text] [doi: 10.2196/jmir.4345] [Medline: 26715363]
22. Lee K, Hoti K, Hughes JD, Emmerton L. Dr Google and the consumer: a qualitative study exploring the navigational needs and online health information-seeking behaviors of consumers with chronic health conditions. J Med Internet Res 2014;16(12):e262 [FREE Full text] [doi: 10.2196/jmir.3706] [Medline: 25470306]
23. Linden K, Berg M, Sparud-Lundin C. Web-based information for pregnant women and new mothers with type 1 diabetes--a description of the development process. BMC Med Inform Decis Mak 2012 Nov 20;12:134 [FREE Full text] [doi: 10.1186/1472-6947-12-134] [Medline: 23167552]
24. Yu CH, Parsons JA, Hall S, Newton D, Jovicic A, Lottridge D, et al. User-centered design of a web-based self-management site for individuals with type 2 diabetes-providing a sense of control and community. BMC Med Inform Decis Mak 2014 Jul 23;14:60 [FREE Full text] [doi: 10.1186/1472-6947-14-60] [Medline: 25056379]
25. Prothero L, Georgopoulou S, de Souza S, Bosworth A, Bearne L, Lempp H. Patient involvement in the development of a handbook for moderate rheumatoid arthritis. Health Expect 2017 Dec;20(2):288-297 [FREE Full text] [doi: 10.1111/hex.12457] [Medline: 27086728]
26. Flexner WA, McLaughlin CP, Littlefield JE. Discovering what the health consumer really wants. Health Care Manage Rev 1977;2(4):43-49. [Medline: 10305184]
27. Gold J, Pedrana AE, Stoove MA, Chang S, Howard S, Asselin J, et al. Developing health promotion interventions on social networking sites: recommendations from The FaceSpace Project. J Med Internet Res 2012 Feb 28;14(1):e30 [FREE Full text] [doi: 10.2196/jmir.1875] [Medline: 22374589]
28. Oakman J, Kinsman N, Briggs AM. Working with persistent pain: an exploration of strategies utilised to stay productive at work. J Occup Rehabil 2017 Mar;27(1):4-14. [doi: 10.1007/s10926-016-9626-5] [Medline: 26800700]
29. Oakman J, Keegel T, Kinsman N, Briggs AM. Persistent musculoskeletal pain and productive employment; a systematic review of interventions. Occup Environ Med 2016 Mar;73(3):206-214. [doi: 10.1136/oemed-2015-103208] [Medline: 26740689]
30. Li L, Sayre EC, Xie H, Falck RS, Best JR, Liu-Ambrose T, et al. Efficacy of a community-based technology-enabled physical activity counseling program for people with knee osteoarthritis: proof-of-concept study. J Med Internet Res 2018 Apr 30;20(4):e159 [FREE Full text] [doi: 10.2196/jmir.8514] [Medline: 29712630]
31. Slater H, Jordan JE, Chua J, Schütze R, Wark JD, Briggs AM. Young people's experiences of persistent musculoskeletal pain, needs, gaps and perceptions about the role of digital technologies to support their co-care: a qualitative study. Br Med J Open 2016 Dec 09;6(12):e014007 [FREE Full text] [doi: 10.1136/bmjopen-2016-014007] [Medline: 27940635]
32. Boote J, Telford R, Cooper C. Consumer involvement in health research: a review and research agenda. Health Policy 2002 Aug;61(2):213-236. [doi: 10.1016/S0168-8510(01)00214-7] [Medline: 12088893]
33. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
34. Macqueen G. Applied Thematic Analysis. Los Angeles: Sage Publications; 2018.
35. Creswell J. Research Design. Thousand Oaks, CA: Sage Publications; 2018.

36. Barbour RS. Checklists for improving rigour in qualitative research: a case of the tail wagging the dog? Br Med J 2001 May 5;322(7294):1115-1117 [FREE Full text] [Medline: 11337448]
37. Fusch PI, Ness LR. Are we there yet? Data saturation in qualitative research. Qual Rep [Internet] 2015;20(9):1408-1416 [FREE Full text]
38. Mason M. Sample size and saturation in PhD studies using qualitative interviews. Forum Qual Soc Res 2010;11(3). [doi: 10.17169/fqs-11.3.1428]
39. Hewison J, Haines A. Overcoming barriers to recruitment in health research. Br Med J 2006 Aug 05;333(7562):300-302 [FREE Full text] [doi: 10.1136/bmj.333.7562.300] [Medline: 16888308]
40. Pang PC, Temple-Smith M, Bellhouse C, Trieu VH, Kiropoulos L, Williams H, et al. Online health seeking behaviours: what information is sought by women experiencing miscarriage? Stud Health Technol Inform 2018;252:118-125. [Medline: 30040693]
41. Pang PC, Chang S, Pearce J, Verspoor K. Online Health Information Seeking Behaviour: Understanding Different Search Approaches. In: Proceedings of the 18th Pacific Asia Conference on Information Systems (PACIS 2014).: AIS; 2014 Presented at: PACIS 2014; 24 Jun 2014; Chengdu, China.
42. Pang PC, Verspoor KM, Chang S, Pearce J. Conceptualising health information seeking behaviours and exploratory search: result of a qualitative study. Health Technol 2015 Feb 10;5(1):45-55. [doi: 10.1007/s12553-015-0096-0]
43. Pang PC, Clavisi O, Chang S. Engaging Consumers with Musculoskeletal Conditions in Health Research: a user-centred perspective. Stud Health Technol Inform 2017;239:104-111. [doi: 10.3233/978-1-61499-783-2-104] [Medline: 28756444]
44. Walsh L, Hill S, Allan M, Balandin S, Georgiou A, Higgins I, et al. A content analysis of the consumer-facing online information about My Health Record: Implications for increasing knowledge and awareness to facilitate uptake and use. Health Inf Manag 2018 Sep;47(3):106-115. [doi: 10.1177/1833358317712200] [Medline: 28589741]
45. Tieu L, Sarkar U, Schillinger D, Ralston JD, Ratanawongsa N, Pasick R, et al. Barriers and facilitators to online portal use among patients and caregivers in a safety net health care system: a qualitative study. J Med Internet Res 2015 Dec 03;17(12):e275 [FREE Full text] [doi: 10.2196/jmir.4847] [Medline: 26681155]
46. Khodadad Hosseini SH, Behboudi L. Brand trust and image: effects on customer satisfaction. Int J Health Care Qual Assur 2017 Aug 14;30(7):580-590. [doi: 10.1108/IJHCQA-04-2016-0054] [Medline: 28809590]
47. Gilligan C, Kypri K, Bourke J. Social networking versus facebook advertising to recruit survey respondents: a quasi-experimental study. JMIR Res Protoc 2014 Sep 17;3(3):e48 [FREE Full text] [doi: 10.2196/resprot.3317] [Medline: 25230740]
48. Kapp JM, Peters C, Oliver DP. Research recruitment using Facebook advertising: big potential, big challenges. J Cancer Educ 2013 Mar;28(1):134-137. [doi: 10.1007/s13187-012-0443-z] [Medline: 23292877]
49. Lohse B, Wamboldt P. Purposive facebook recruitment endows cost-effective nutrition education program evaluation. JMIR Res Protoc 2013 Aug 15;2(2):e27 [FREE Full text] [doi: 10.2196/resprot.2713] [Medline: 23948573]
50. Pew Research Center. 2018. Social Media Use in 2018 URL: http://assets.pewresearch.org/wp-content/uploads/sites/14/2018/03/01105133/PI_2018.03.01_Social-Media_FINAL.pdf [WebCite Cache ID 72lSPGaoo]
51. Vraga E, Bode L, Troller-Renfree S. Beyond self-reports: using eye tracking to measure topic and style differences in attention to social media content. Commun Methods Meas 2016 Apr 20;10(2-3):149-164. [doi: 10.1080/19312458.2016.1150443]
52. Huang Y, Huo S, Yao Y, Chao N, Wang Y, Grygiel J, et al. Municipal Police Departments on Facebook: What Are They Posting and Are People Engaging? In: Proceedings of the 17th International Digital Government Research Conference on Digital Government Research. USA: ACM; 2016 Presented at: The 17th International Digital Government Research Conference on Digital Government Research; 2016; New York, NY p. 366-374. [doi: 10.1145/2912160.2912189]
53. Kite J, Foley BC, Grunseit AC, Freeman B. Please Like Me: Facebook and public health communication. PLoS One 2016;11(9):e0162765 [FREE Full text] [doi: 10.1371/journal.pone.0162765] [Medline: 27632172]
54. Freeman B, Potente S, Rock V, McIver J. Social media campaigns that make a difference: what can public health learn from the corporate sector and other social change marketers? Public Health Res Pract 2015 Mar 30;25(2):e2521517 [FREE Full text] [doi: 10.17061/phrp2521517] [Medline: 25848735]
55. Aggarwal D, Zhang W, Hoang T, Ploderer B, Vetere F, Bradford M. SoPhy: A Wearable Technology for Lower Limb Assessment in Video Consultations of Physiotherapy. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. New York, NY: ACM; 2017 Presented at: CHI 2017; 2017; N/A p. 3916-3928. [doi: 10.1145/3025453.3025489]
56. Kaye J, Whitley EA, Lund D, Morrison M, Teare H, Melham K. Dynamic consent: a patient interface for twenty-first century research networks. Eur J Hum Genet 2015 Feb;23(2):141-146 [FREE Full text] [doi: 10.1038/ejhg.2014.71] [Medline: 24801761]
57. 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 Jan 13;3(1):e3 [FREE Full text] [doi: 10.2196/medinform.3525] [Medline: 25586934]

58. Furberg RD, Ortiz AM, Moultrie RR, Raspa M, Wheeler AC, McCormack LA, et al. A digital decision support tool to enhance decisional capacity for clinical trial consent: design and development. JMIR Res Protoc 2018 Jun 06;7(6):e10525 [FREE Full text] [doi: 10.2196/10525] [Medline: 29875084]
59. Arthritis and Osteoporosis Victoria. A problem worth solving: the rising cost of musculoskeletal conditions in Australia. Elsternwick, VIC, Australia: Arthritis and Osteoporosis Victoria; 2013.

Abbreviations

CN: consumer needs
MSK: musculoskeletal
RS: research strategies
RQ: research question
SNSs: social networking sites

Edited by G Eysenbach; submitted 02.09.18; peer-reviewed by A Cyr, L Walsh; comments to author 21.09.18; revised version received 08.10.18; accepted 08.10.18; published 30.10.18.

Please cite as:
Pang PCI, Chang S, Verspoor K, Clavisi O
The Use of Web-Based Technologies in Health Research Participation: Qualitative Study of Consumer and Researcher Experiences
J Med Internet Res 2018;20(10):e12094
URL: https://www.jmir.org/2018/10/e12094/
doi:10.2196/12094
PMID:30377139

Original Paper

The Impact of Disease-Modifying Therapy Access Barriers on People With Multiple Sclerosis: Mixed-Methods Study

Kristina F Simacek[1], MA, PhD; John J Ko[2], MS, PharmD; Debbie Moreton[3], BS; Stefan Varga[4], PharmD; Kristen Johnson[2], MPH, PhD; Bozena J Katic[1], MPH, PhD

[1]Research and Development, PatientsLikeMe, Cambridge, MA, United States
[2]Health Economic and Outcomes Research, Novartis Pharmaceuticals Corporation, East Hanover, NJ, United States
[3]PatientsLikeMe, Cambridge, MA, United States
[4]Jefferson College of Population Health, Thomas Jefferson University, Philadelphia, PA, United States

Corresponding Author:
Kristina F Simacek, MA, PhD
Research and Development
PatientsLikeMe
160 2nd Street
Cambridge, MA, 02142
United States
Phone: 1 617 499 4003
Email: ksimacek@patientslikeme.com

Abstract

Background: In the United States, people with relapsing-remitting multiple sclerosis (RRMS) can face difficulty accessing disease-modifying therapies (DMTs) because of insurance, pharmacy, or provider policies. These barriers have been associated with poor adherence and negative health outcomes.

Objective: The goals of this study were to describe the overall occurrence of difficulties and delays associated with gaining access to DMTs among people with RRMS, to assess DMT adherence during periods of reduced access, and to contextualize the patients' journey from receipt of a prescription for DMT to obtaining and taking their medication when faced with access barriers.

Methods: We recruited US-based adults self-reporting RRMS from a Web-based health data-sharing social network, PatientsLikeMe. Individuals were invited to complete a Web-based survey if they reported a diagnosis of RRMS and were prescribed a DMT for MS. Follow-up phone interviews were conducted with 10 respondents who reported experiencing an MS-related relapse during the time they had experienced challenges accessing DMTs.

Results: Among 507 survey completers, nearly half were either currently experiencing an issue related to DMT assess or had difficulty accessing a DMT in the past (233/507, 46.0%). The most frequently reported reasons for access difficulty were authorization requirements by insurance companies (past issues: 78/182, 42.9%; current issues: 9/42, 21%) and high out-of-pocket costs (past issues: 54/182, 29.7%; current issues: 13/42, 31%). About half (20/39, 51%) of participants with current access issues and over a third (68/165, 41.2%) of those with past issues went without their medication until they could access their prescribed DMT. Relapses were reported during periods of reduced DMT access for almost half (56/118, 47.5%) of those with past issues and nearly half (22/45, 49%) of those with current issues. Resolving access issues involved multiple stakeholder agents often coordinated in a patient-led effort. Among those who had resolved issues, about half (57/119, 47.9%) reported that doctors or office staff were involved, under half (48/119, 40.3%) were involved themselves, and about a third (39/119, 32.8%) reported the drug manufacturer was involved in resolving the issue. Follow-up interviews revealed that the financial burden associated with obtaining a prescribed DMT led to nonadherence. Additionally, participants felt that DMT treatment delays and stress associated with obtaining the DMT triggered relapses or worsened their MS.

Conclusions: This study expands current research by using a patient-centered, mixed-methods approach to describe barriers to MS treatment, the process to resolve barriers, and the perceived impact of treatment barriers on outcomes. Issues related to DMT access occur frequently, with individuals often serving as their own agents when navigating access difficulties to obtain their medication(s). Support for resolution of DMT access is needed to prevent undue stress and nonadherence.

(J Med Internet Res 2018;20(10):e11168) doi:10.2196/11168

KEYWORDS

cost sharing; insurance; mixed methods; multiple sclerosis; out-of-pocket costs; patient adherence; pharmaceutical services; self-report; surveys and questionnaires

Introduction

Multiple sclerosis (MS) is a chronic autoimmune disease of the central nervous system affecting an estimated 450,000-550,000 people in the United States [1,2]. The financial burden accompanying MS is a central component of the disease experience, ranking second among all chronic conditions in direct costs behind congestive heart failure [3]. In addition to the impact of decreased productivity and unemployment [4-6], patients often shoulder the burden of high out-of-pocket costs for medications, tests, magnetic resonance imaging (MRI), medical equipment, and inpatient or outpatient visits [3,7,8].

Disease-modifying therapies (DMTs) slow the disease progression and related disability and are the prevailing treatment for people with MS [9]. Responding to rising costs of DMTs [10,11], insurers have adopted strategies to optimize the utilization of specialty medications through management, including requiring prior authorization, step therapy (where a patient must fail a preferred medication first), or tiered formularies that increase cost sharing for patients [12,13]. Falling into the “specialty drug” tier of most benefits plans, DMTs are subject to higher out-of-pocket costs (ie, tiers that require high coinsurance or copayment) than nonspecialty drugs [14]. Due to the clinical benefits of specialty drugs like DMTs, their use remains relatively insensitive to such cost-sharing programs [14,15]. For people with MS, high out-of-pocket costs for DMTs have been associated with lower adherence and inappropriate disease management, which leaves patients at increased risk of relapse and payers vulnerable to increased associated resource use [16,17].

A growing body of research has explored the impact of current insurance coverage and related barriers to DMT access on people with MS. An estimated 25% of people with MS taking DMTs receive these drugs at little to no cost through pharmaceutical company financial aid programs, and such programs have been shown to increase the adherence [18,19]. Discounted drug programs can be essential in mitigating prohibitive financial barriers for patients; however, regulations may prevent those with government-based insurance coverage from using these programs [20], so they may not be an ideal solution for all patients.

One understudied aspect of access to MS DMTs is the patient experience in navigating the treatment acquisition landscape. Although prior research has focused on the number of people receiving DMTs, how they paid for their treatment, and insurance-related access barriers [19], it is not yet clear how people with MS perceive and navigate the current health care system to obtain their medication and how they perceive this impacts their health outcomes and self-reported quality of life.

The goals of this study are to describe the overall occurrence of difficulties and delays associated with gaining access to DMTs among people with MS, to assess the impact of access barriers on DMT medication adherence and MS outcomes, and to contextualize the experience of obtaining a DMT after the receipt of a prescription.

Methods

Study Design

This mixed-methods study included a population of people with MS who were members of PatientsLikeMe as of February 2016. PatientsLikeMe is a real-time, data-sharing, and research platform of patient communities devoted to a variety of life-changing health conditions. As of July 1, 2018, PatientsLikeMe had >600,000 members with >2800 different health conditions, including 59,853 members with MS, and among those reporting their variant of MS, most (32,732/50,868, 64.35%) reported having relapsing-remitting MS (RRMS). Of note, participants were not compensated. This study was exempted from review by the New England Independent Review Board on February 26, 2016 (study #16-082).

The study was designed into 2 distinct phases, consisting of a Web-based survey and qualitative interviews following the survey completion with a subset of survey participants who had experienced a relapse during their access issue.

Phase 1: Web-Based Survey

A sample of people with RRMS from the PatientsLikeMe community participated in a cross-sectional, Web-based questionnaire fielded by PatientsLikeMe [21]. A closed design was used, meaning only members of PatientsLikeMe who were invited to take part and who had logged in with their password could see the survey. An informed consent document was displayed prior to starting the survey, and a waiver of documentation of informed consent was obtained from the New England Independent Review Board. The survey was administered electronically for 4 weeks in February and March 2016.

Survey Development

The survey included demographics and a series of questions concerning experiences with access to DMTs. We used 3 mutually exclusive branching options to segment respondents who (1) were currently experiencing a DMT access barrier, (2) had past (but not current) barrier accessing DMTs, and (3) never experienced an access barrier to their DMTs. Preliminary survey items were derived from targeted literature review and consultation with MS experts; a PatientsLikeMe patient member (DM) provided feedback on items, survey length, and participant interpretability. Prior to fielding the survey, it went through a series of reviews among investigators and was electronically pretested for design elements, question ordering, and flow; see Multimedia Appendix 1 for survey questions.

Survey Population

Eligible participants were aged ≥18 years, who self-reported residence within the United States, reported a physician diagnosis of RRMS, and had a recent (90-day) activity on the

PatientsLikeMe website. Eligible participants were sent a private survey invitation through the PatientsLikeMe website. Notably, participants were screened out if they did not confirm the above eligibility criteria or reported never being prescribed a DMT for RRMS.

Measures

DMT access barriers included any of the following: lack of insurance; insurance not covering the DMT; prior authorization documentation requirement; high out-of-pocket costs; requirement to take another DMT before the prescribed DMT (ie, step through); inability to obtain the DMT at their desired pharmacy or infusion center; or other difficulties.

Adherence was measured by asking participants how often they take their medication as prescribed. Responses could be given in 10% increments from 0% (none of the time) to 100% (always).

Health literacy was measured using respondents' level of confidence filling out medical forms by themselves [22,23]. In addition, a dichotomous variable was created to classify those who were "quite a bit" or "extremely" confident as having high health literacy; those who responded "somewhat," "a little," or "not at all" as having low health literacy.

Stress was measured on a 10-point numeric scale. Respondents who had experienced an access barrier were asked how stressful their most recent DMT access issue was from 0 (not at all) to 10 (greatly). Those who had not experienced an access issue were instead asked about their stress level in the last 7 days.

In addition to survey responses, most recent Patient-Determined Disease Steps (PDDS) score [24-27] was obtained from participants' PatientsLikeMe profiles. PDDS is measured on a 9-point scale from 0 (Normal) to 8 (Bedridden).

Analyses

In this study, subgroups (current, past, or never had DMT issue) were defined by the following 2 questions: "Which option best describes your experience with accessing or receiving your DMT medication for MS?" (current DMT access difficulty, a past difficulty, or never had a difficulty) and "Have you ever had any of the following difficulties accessing or receiving DMT medications for MS?" (a select-all-that-apply list including items such as covered by insurance plan and it required authorizing documentation).

During analysis, it was discovered that some participants reported they had never experienced a DMT access issue; however, when queried about specific access issues, they selected a specific DMT access barrier (eg, authorizing documentation, high out-of-pocket costs, medications not covered by insurance). For clarification, a 2-question follow-up survey was fielded in June and July 2016 to 108 respondents who fell into this category.

The first question was closed-ended: "Were any of the following situations burdensome enough to delay or prevent you from getting your DMT medication? (check all that apply)." Answer options included (1) having to fill out paperwork or get other documentation for DMT access or (2) DMT out-of-pocket costs too high. The second question was open-ended: "Please tell us a little bit more about this difficulty and the effect it had on you (if any)."

Variables with closed-ended response options were analyzed using descriptive and summary statistics. We used Wilcoxon rank-sum tests to compare groups with nonnormally distributed values; *t* tests were used to compare groups with normally distributed values. An alpha value of ≤.05 was considered significant. Statistical analyses were performed using SAS software, version 9.4 (Cary, NC, USA). Open-ended questions were coded for themes using conventional content analysis in ATLAS.ti software (ATLAS.ti Scientific Software Development GmbH, Berlin, Germany, version 5.0) [28,29].

Data Exclusion

A validated sample of survey respondents excluded 121 respondents from the analyzed sample who gave conflicting responses to the survey branches. The total number of respondents in the "never" category reduced from the full sample of 507 respondents to 386.

Phase 2: Qualitative Interviews

Interview methods are described using the consolidated criteria for reporting qualitative research [30]. During the second phase of this study, 10 respondents who completed the questionnaire were selected to participate in a single, 60-minute follow-up interview between April and June 2016. Interviews were conducted at the participants' convenience via phone or videoconference by researchers (KFS and BJK). The interviewers had no prior relationship with interviewees. Interviews were audiorecorded and transcribed. Participants were not given transcripts or findings for review.

Interview Sample Population

Participants were selected for interviews based on their survey response, reporting either a current or past issue with DMT access, at least one MS relapse during the period of their DMT access difficulty, and consent to a follow-up interview in the first survey.

Interview Guide Development

The goal of the interview was to supplement information collected in the questionnaire, among participants who had experienced DMT access barriers and a negative health event (eg, MS relapse) during the assess issue, using qualitative data (see Multimedia Appendix 2 for the Interview Guide). The following 4 research questions directed the interview guide: (1) How did people resolve their DMT access issues? (2) What are the clinical consequences of DMT access barriers? (3) What are the impacts of DMT access issues on quality of life? and (4) Who are the agents involved in helping people access their DMT?

Analyses

Interviews were analyzed for themes by researchers using a constant comparative method [31,32]. To represent interactions between study participants and the agents or organizations involved in resolving their access to DMTs, an aggregate egocentric social network map was created based on the interviews [33,34]. To create the map, an iterative process was used to draw connections between each interviewee and others involved in DMT access resolution. Each subsequent interviewee was compared with the previous ones until all connections mentioned in the interviews were placed on the generalized map, with the “ego” (study participant) in the center.

Results

Survey Results

Survey Sample

Of 5239 people invited to the survey, 947 viewed the invitation and 584 participated in the survey, for a participation rate of 61.6%. A total of 507 participants completed the survey, for a completion rate of 86.8% (507/584). A subset of 40 participants (out of 108 who were invited) completed the follow-up survey, for a participation rate of 85%). Among survey completers, 78.3% (397/507) were females, 89.6% (441/507) were white, and participants were, on average, 49 years old (Table 1). Of the survey completers, 66% (234/507) had high health literacy; however, the percentage of those with high health literacy was lowest among those currently experiencing access issues (25/45, 56%, of those with a current issue vs 138/188, 73.4%, of those with a past issue and 127/153, 83.0%, of those who never had issues had high health literacy). This difference was statistically significant (χ^2_1=14.6, P<.001). Median PDDS score was 4 (range, 1-8; interquartile range [IQR], 3.0), indicating some gait disability. Of those in the validated sample, 60.4% (233/386) had experienced difficulty accessing a DMT in the past (188/386, 48.7%) or were currently experiencing an issue related to DMTs access (45/386, 11.7%). Average time to receive the originally prescribed DMT after experiencing an access barrier was 8.3 (SD 16.5) weeks.

Reasons for Disease-Modifying Therapy Access Difficulties

The most frequently reported reasons for DMT-related access difficulties were “insurance required authorizing documentation” (9/42, 21.4%, current issue and 78/182, 42.9%, past issue) and “high out-of-pocket costs” (13/42, 31.0%, current issue and 54/182, 29.7%, past issue; Table 2). Other reasons included administrative coordination problems among insurance companies, pharmacies, and clinician’s offices.

Among the 40 follow-up respondents, 60% (24/40) clarified that the DMT access barrier delayed or prevented them from getting their DMT medication. The reasons for access problems mirrored the results obtained from the full survey, with about half of those who had experienced barriers attributing them to “authorizing documentation” (58%, 14/24), a third to “high out-of-pocket costs” (33%, 8/24), and several to “not covered by insurance” (8%, 2/24). Furthermore, a slight majority of the 40 follow-up respondents (66%, 16/24) confirmed that their access difficulty did not delay or prevent them from getting their DMT.

Adherence to Prescribed Disease-Modifying Therapies During Periods of Decreased Access

Respondents frequently went without any RRMS medication until they could obtain their prescribed DMT. Nearly half (68/165, 41.2%) of respondents who had experienced a past access issue reported going without their medication(s) until they could access their prescribed DMT, 4.8% (8/165) switched to a different DMT, 1.8% (3/165) continued their old medication, and 1.2% (2/165) received a different DMT until receiving the originally prescribed DMT. Among respondents currently experiencing a DMT access issue, about half (20/39, 51%) responded they were not currently taking any medications for their MS, 23% (13/39) continued their old medication, 13% (5/39) took a newly prescribed DMT, and 3% (1/39) reported they were instead taking a newly prescribed non-DMT medication for MS (Table 3).

Self-reported adherence to DMT medication during an access barrier (mean 8.97 [SD 2.47]) was significantly lower than self-reported typical DMT adherence (mean 9.61 [SD 1.0]). A paired t test showed that this difference was statistically significant (t_{101}=−2.48, P=.02).

Outcomes During Periods of Decreased Disease-Modifying Therapy Access: Stress and Multiple Sclerosis Relapse

Among respondents who experienced a DMT access barrier, 49% (22/45) of those with a current access issue reported at least one MS relapse during the time of the barrier; 29.8% (56/188) self-reported at least one MS relapse during a past DMT access issue. The Wilcoxon rank-sum tests showed significantly higher stress levels among those who experienced at least one MS relapse during a past access issue (n=48; median, 8.5; IQR, 7.0-10.0) than among those who did not experience relapse (n=113; median, 7.0; IQR, 5-9; Z=3.228, P=.001); this effect did not reach significance for those with a current issue (n=19; median, 10.0; IQR, 7.0-10.0 vs n=20; median 7.0; IQR, 4.3-8.8; Z=−1.835, P=.08).

Stakeholder Agents Involved in Disease-Modifying Therapy Access

Among respondents who experienced past difficulties gaining access to DMTs, 47.9% (57/119) involved doctors or office staff to help resolve the DMT access issue and 40.3% (48/119) said they were at least partially responsible for resolving the issue themselves. The remaining agents involved in resolving the issue were the drug manufacturers (39/119, 32.8%), pharmacy or specialty pharmacy (31/119, 26.1%), insurance companies (26/119, 21.8%), and infusion centers (6/119, 5.0%). Few caregivers were involved in resolving the access barriers (2/119, 1.7%).

Table 1. Participant characteristics.

Characteristic	Total completed (N=507)	Validated sample[a] (n=386)	Current issue (n=45)	Past issue (n=188)	Never had issue (n=153)
Age (years), mean (SD)	49.1 (10.4)	49.7 (10.2)	50.2 (9.7)	48.9 (10.8)	50.5 (9.7)
Female, n (%)	397 (78.3)	308 (79.8)	39 (86.7)	154 (81.9)	115 (75.2)
White, n (%)	441 (89.6)	335 (89.8)	40 (90.9)	161 (88.5)	134 (91.2)
Ethnicity, n (%)					
Not Hispanic	462 (94.7)	357 (95.7)	41 (93.2)	173 (95.6)	143 (96.6)
Number of comorbidities, median (range, IQR[b])	1 (1-27, 1.0)	1 (1-27, 2.0)	1 (1-21, 3.0)	1 (1-27, 1.0)	1 (1-15, 1.0)
Education[c], n (%)					
High school or less	53 (12.7)	39 (12.2)	6 (16.2)	21 (13.6)	12 (9.4)
Some college	165 (39.6)	126 (39.4)	18 (48.6)	56 (36.1)	52 (40.6)
College degree	131 (31.4)	107 (33.4)	5 (13.5)	57 (36.8)	45 (35.2)
Postgraduate work	68 (16.3)	48 (15.0)	8 (21.6)	21 (13.6)	19 (14.8)
High Health Literacy[d], n (%)					
Quite a bit or extremely	234 (66.1)	182 (65.5)	25 (55.5)	138 (73.4)	127 (83.0)
Health Insurance Status[e], n (%)					
Employer based	209 (47.5)	162 (47)	18 (43.9)	76 (45.5)	68 (49.6)
Direct	34 (7.7)	26 (7.5)	2 (4.9)	14 (8.4)	10 (7.3)
Medicare	131 (29.8)	103 (29.9)	10 (24.4)	56 (33.5)	37 (27.0)
Medicaid	38 (8.6)	32 (9.3)	8 (19.5)	11 (6.6)	13 (9.5)
Military	7 (1.6)	7 (2.0)	1 (2.4)	3 (1.8)	3 (2.2)
Veterans Affairs	9 (2.1)	8 (2.3)	1 (2.4)	2 (1.2)	5 (3.6)
None	8 (1.8)	4 (1.2)	1 (2.4)	3 (1.8)	0 (0)
Other	4 (0.8)	3 (0.8)	0 (0)	2 (1.2)	1 (0.7)
PDDS[f] score, median (range, IQR)	4 (1-8, 3.0)	4 (1-8, 3.0)	4 (1-8, 3.0)	4 (1-8, 3.0)	4 (1-8, 4.0)
Had relapse during access issue (self-reported), n (%)	N/A[g]	N/A	22 (48.9)	56 (29.8)	N/A
Average delay in weeks[h,i], mean (SD)	N/A	N/A	N/A	8 (16.5)	N/A

[a] Subgroups (current, past, or never had a disease-modifying therapy [DMT] issue) were drawn from a validated sample based on the questions: “Which option best describes your experience with accessing or receiving your DMT medication for MS?” “Never” includes only those who selected none of the DMT access issues and “never” to “Which option best describes your experience with accessing or receiving your DMT medication for MS?”

[b] IQR: interquartile range.

[c] Validated sample, n=320

[d] Validated sample, n=278

[e] Validated sample, n=345

[f] PDDS: Patient Determined Disease Steps.

[g] N/A: not applicable.

[h] Asked only of those who eventually received the originally prescribed DMT.

[i] Past issue, n=64.

Table 2. Reported reasons for the disease-modifying therapy (DMT) access issue.

Source of access issue[a]	Past issue[b], n (%)	Current issue[c], n (%)	Follow-up[d], n (%)
Insurance required authorizing documentation	78 (42.9)	9 (21.4)	14 (58.3)
High out-of-pocket costs	54 (29.7)	13 (31.0)	8 (33.3)
Not covered by my insurance plan	20 (11.0)	8 (19.0)	2 (8.3)
I do not have insurance	17 (9.3)	3 (7.1)	N/A[e]
Not at my desired pharmacy or infusion	14 (7.7)	2 (4.8)	N/A
Required to take one additional DMT	9 (4.9)	6 (14.3)	N/A
Other[f,g]	37 (20.3)	8 (19.0)	N/A
I don’t know	12 (6.6)	6 (14.3)	N/A

[a] Among a validated sample of those with past or current difficulties, access reasons were only asked of those who received a DMT prescription from their doctor. Question text for current issue was: “What difficulties are you having accessing or retrieving your DMT medication? Check all that apply”; question text for past access issue was: “Thinking about your most recent MS DMT access issue, what difficulties did you have accessing or receiving your DMT medication? Check all that apply.”

[b] Past issue, n=182.

[c] Current issue, n=42.

[d] Follow-up, n=24.

[e] N/A: not applicable.

[f] Other past reasons included administrative problems (n=9), provider changed or could not authorize (n=5), insurance or pharmacy denied drug or changed policies (n=5), insurance status change (n=4), appointment or prescription delay by the provider (n=4), paperwork issue (n=2), and other (n=5).

[g] Other current reasons included insurance policy changes or coverage loss (n=2), doctor or hospital problems (n=2), administrative problems (n=1), and switched drug (n=1).

Table 3. Medication status during past and current access issue.

Medication status[a]	Past issue[b], n (%)	Current issue[c], n (%)
Not taking any medication or went without medication	68 (41.2)	20 (51.3)
I received my disease-modifying therapy (DMT) medication within a reasonable amount of time,	56 (33.9)	N/A[d]
I was prescribed a new DMT instead	8 (4.8)	5 (12.8)
I continued taking my old medication	3 (1.8)	13 (33.4)
Received another DMT before receiving my prescribed DMT	2 (1.2)	N/A
I am taking a newly prescribed other non-DMT medication	N/A	1 (2.6)
Other	28 (17.0)	N/A

[a] Asked of respondents who had insurance or did not answer that difficulty obtaining DMT at a pharmacy or infusion center was their primary DMT access reason. Question text for current access status was: “What other MS medication(s) are you taking while your DMT medication access issue is being resolved?”; question text for past access issue was “Pick the option that best describes how your most recent MS DMT access issue was resolved.”

[b] Past issue, n=165.

[c] Current issue, n=39.

[d] N/A: not applicable.

Qualitative Interviews

Interview Sample

Qualitative interviews were conducted among 10 survey respondents who experienced at least one MS relapse during a past or current period of decreased access to DMTs. Participants were predominantly females (9/10, 90%) with the mean age of 54 (range, 42-64) years; of them, 5 reported being on Medicaid or Medicare, 3 on employer-sponsored insurance, 1 did not specify insurance type, and 1 had no insurance. The median duration of MS was 9.5 (range, 2-15) years.

Themes From Qualitative Interviews

We identified several themes in the interviews; detailed examples are shown in Table 4.

Theme 1: Financial Burden Begins Prior to the Disease-Modifying Therapy Access Barrier and Can Impact Adherence to the Therapy

Many participants reported that prior to being prescribed a DMT, they had been in financial distress due to MS diagnostic costs and/or loss of income due to inability to work because of their symptoms (Table 4). For example, one participant said she had spent “all our savings” (Female, age 58 years) on medical costs related to her MS diagnosis, such as MRI tests. Another filed for bankruptcy because of bills stemming from his initial MS diagnosis. Most interviewees had stopped or reduced paid work because of MS symptoms, and over half (6/10, 60%) sought or received disability pay or subsisted on a fixed monthly income lower than the amount of their monthly copayment for their DMT medication. Most interviewees (8/10, 80%) went without DMT medication during their access barrier; the remaining 2 took a DMT they had been prescribed in the past until the access barrier was resolved.

Theme 2: Disease-Modifying Therapy Access Barriers are Associated With Stress and Relapses

Many participants felt delays in DMT treatment, and the stress associated with the process of obtaining the DMT triggered relapses or worsened their MS. As one participant who experienced a combination of billing errors and finding her DMT out of stock at her infusion center asked, “Why do *I* have to deal with this crap? You know how they say that stress makes it worse?” (Female, age 56 years). Several participants experienced worsening fatigue and cognitive problems related to MS and the lack of DMT, making efforts to resolve their access issue more difficult.

Theme 3: Disease-Modifying Therapy Access Issues Affect Quality of Life

Access difficulties impacted multiple facets of participants’ lives; emotional and interpersonal impacts were commonly mentioned. Emotional impacts included situational problems like “frayed nerves” (female, age 58 years) and exacerbation of pre-existing mental health comorbidities, such as depression. In addition to the time and effort spent on trying to access a DMT, several participants reported that the uncertainty of having unstable health because of going without a DMT made it difficult to schedule social events in advance.

Theme 4: Personal Resources Enable Access to Overcome Disease-Modifying Therapy Access Barriers

Several participants who succeeded in obtaining their DMT reported that they leveraged knowledge and skills from working in medical billing and other health care areas to expedite the process. This high level of health care literacy obtained through work experience facilitated their ability to document the issue and reach the appropriate agents who could help resolve the problem quickly.

Table 4. Themes arising from participant interviews.

Theme	Example quotes
Financial burden begins prior access issue and impacts the disease-modifying therapy (DMT) adherence	*I had to declare bankruptcy because of my first doctor’s bill. We accumulated US $15,000 in debt through the MRIs because they only cover half of one MRI per year and I had 6 that first year…I went from making US $6500 a month to US $1400 a month [on disability] with two kids and a vehicle. I got a little part time job that’s 12 hours a week, US $10 an hour. That pays for my drugs.* [Male, age 42 years]
	Patients are not getting help. We cannot afford insurance. My discretionary income is US $10, that’s why I needed a physician who accepted cash. With co-pay, deductible and premium I could not afford that. [Female, age 45 years]
	I couldn’t afford the co pay and just quit taking the DMM [disease modifying medication]. I also quit taking other medications I could not afford to purchase. [Female, age 64 years]
DMT access problems and related stress leads to multiple sclerosis relapses	*I’ve had a series of bad attacks when the prescription lapsed and when the insurance lapsed. I have some severe damage where it comes to process from the printed page and to spit it back out again. That ability is gone unless it’s in context…I had lesions confirmed. The area with the vocabulary. I’m also a bit slow on the processing. There are things I don’t do so well. It takes me that little extra moment, so there’s this pause in my conversation…Those are the two areas of the brain affected.* [Female, age 58 years]
	You spend 45 min fighting on the phone it’s like working 8 hours. I have to take 2-3 hours of the day for a nap to get my energy back…I’ve done nothing but fight with [my insurance company]. [Female, age 51 years]
	I had a relapse while waiting to get on [DMT]. My left arm is numb and tingling constantly. [Female, age 54 years]
DMT access issues affect the quality of life	*When I didn’t have the medication, I have depression, and that’s not a surprise when you have MS, and it affected me really bad, especially without the [DMT] and I can’t afford it. I kept thinking I don’t know what’s happening inside of me and are things happening to me that I won’t be able to come back from? Not having the medication really affected me emotionally as well.”* [Female, age 59 years, on Medicare with too high co-pay]
	My husband has to work 10 hour days and he’s stressed because of me. I worry about him. If I couldn’t get coverage on HC.gov, I could have gotten on my husband’s plan but that would have been more money, more money out of his paycheck. I went for the lesser of two evils but it’s still US $352 per month. There’s gotta be an in-between. I look for miracles. Someone that doesn’t qualify for disability there has to be some safety net. [Female, age 51 years]
Personal resources enable access to overcome DMT barrier	*I had a 2-month delay to get the authorization. The prescription took over 30 days. First, they sent it to a retail pharmacy and it was rejected. Then it was sent to the wrong specialty pharmacy, and they denied it. Finally, after a month, I got the prescription. I used to do billing for Medicare, so I was familiar with pharmacy denials. Someone else would be lost, they would have had to do without their medicine. The fact that I got the prior authorization, I knew I didn’t have to pay full-price. I knew it should be covered under my plan.* [Female, age 64 years]

Egocentric Social Network of Disease-Modifying Therapy Access

Participants contacted numerous stakeholders during the process of obtaining a DMT. The complexity and intensity of work involved in resolving the problem placed a high burden on them. As one participant described, “I do all of the legwork” (Female, age 58 years). All interviewees contacted their insurance company, physicians, and specialty pharmacy while trying to obtain their DMT. Additional agents contacted included advocacy groups, pharmaceutical companies, government agencies, and hospitals. Figure 1 depicts a conceptual network visualization of these agents as a social network diagram.

Insurance-related access problems were attributed to changes in plans (eg, from an employer-sponsored plan to Medicare), formulary changes by insurance companies’ pharmacy benefit manager, or copayment payment policies. For example, a change to the pharmacy benefit manager contracted with one woman’s insurance company left her with different coverages, new step therapy requirements, and without access to her DMT for months at a time. In addition, insurance changes led to uncertainty about future access; for example, one participant could not find information about whether the Medicaid plan he would enter later in the year would cover his DMT. Finally, some participants could not afford to meet insurance requirements to pay the full price of DMT upfront and wait to be reimbursed later for the portion covered by the insurer.

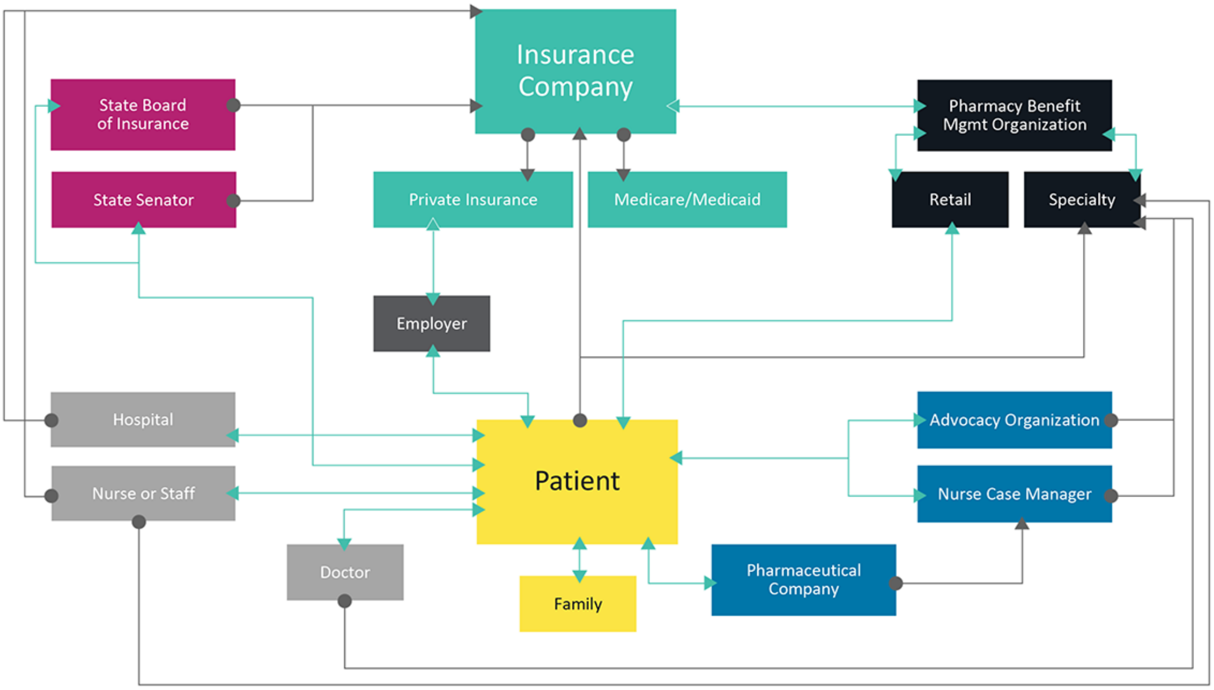
Providers facilitated the documentation of proof of the medical need for advocacy groups or pharmaceutical company programs offering copayment assistance. For some, physician documentation was easily accomplished, while for others, it was a frustrating impediment that required additional calls or visits, and in some cases, delayed treatment access.

Barriers at pharmacies centered around finding a specialty pharmacy to work with their insurance and/or copayment assistance program. Resolving these issues could involve weeks of phone calls and research into plan coverages at specific pharmacies. Several participants reported that insurance companies were unable to provide accurate information about alternative pharmacies that might cover their DMT, forcing them to independently seek this information on the internet or elsewhere.

Nearly all interview participants had sought financial assistance from an advocacy group or a pharmaceutical company; this process was marked by complexity and logistical challenges requiring substantial investment of time and effort to prove need and coordinate stakeholders. Participants reported a range of experiences with patient advocacy groups, from helpful to challenging. Some reported no problems with obtaining copayment assistance from advocacy groups but experienced problems elsewhere in the access process. Most described the advocacy group funding assistance as a “grant” distributed for a certain total amount of money, after which point they would need to reapply. However, these organizations sometimes lacked sufficient funds to (re-)distribute. Others described spending dozens of hours on calls and paperwork to prove eligibility, with one woman reporting her income was deemed US $100 too high for assistance, despite living in an expensive area with a relatively modest income.

Pharmaceutical companies provided participants with copayment assistance or direct access to a DMT. Some participants had difficulty navigating the administrative paperwork necessary to access their copayment assistance programs, which could require original signatures that some found difficult to obtain.

Figure 1. The aggregate egocentric social network of disease-modifying therapy access resolution based on 10 people with multiple sclerosis who experienced a relapse during their access issue.



Others reported that pharmaceutical companies' copayment assistance programs were helpful facilitators, coordinating copayment assistance among patients, providers, and pharmacies or in one instance directly sending medication to a patient when the insurance company and the pharmacy failed to work with the pharmaceutical company's copayment assistance program.

Two interviewees contacted government agencies to enforce the insurance coverage of DMTs that were not being followed; this required additional effort and knowledge to access services to enforce coverage of DMTs.

Finally, participants reported that their families provided emotional and financial support, working to provide insurance and encouraging them to keep trying until they obtained their DMT. However, several reported they felt they were a burden to their spouse or other family members.

Discussion

Principal Findings

In the United States, people with MS can face difficulty accessing DMTs because of insurance, pharmacy, or provider policies. This study demonstrated that issues related to DMT access occur frequently, commonly because of the need for authorizing the documentation, high out-of-pocket costs, and agency or provider coordination problems. Furthermore, participants reported that the effort to overcome barriers could be exceptionally time consuming, complex, and stressful for people with RRMS. This effort required contacting multiple stakeholders in their care administration, including providers, insurers, patient advocacy groups, and others. Furthermore, owing to the lack of DMT, people may navigate this complex process while experiencing disease progression and worsening symptoms. Some participants reported experiencing negative health outcomes during this lack of access, including relapse.

This work represents one of the few studies to both detail the frequency of DMT access issues as well as highlights the patient perspective throughout the DMT acquisition process and the impact this may have on their health outcomes. Furthermore, this study uses a novel approach, blending quantitative and qualitative methods to illuminate the patient experience with DMT access barriers from their own perspective. This approach offers depth and real-world insight that cannot be observed from administrative sources such as claims databases.

Comparison With Prior Work

These findings confirm previous research showing that insurance-related access barriers can be associated with adverse outcomes, such as suboptimal adherence, which is associated with higher medical costs [35-41]. Similar to other studies with people with MS, participants in this study reported that before the DMT issue, many had to reduce or stop working because of functional cognitive decline related to MS [4-6,42,43]. Noting the burden of cost, paperwork, and benefit changes on people with MS trying to obtain DMT, medication advocacy organizations and provider groups have called for system-wide transparency, lower drug prices for DMTs, and policy reforms to assist people with MS with the cost burden of their care [8,44,45].

While the Affordable Care Act of 2010 was implemented to benefit patients by improving the overall health insurance access [46], many patients still lack access to DMTs because of their high cost and specialty status [14]. Cost-sharing efforts on the part of payers have forced many patients to seek financial assistance to defray the costs of DMTs [19,47]. This study showed that even when these programs are available, the logistics of taking part are complex, burdensome, and sometimes unsuccessful, leading to elevated stress levels and, potentially, relapse.

Limitations

There are several limitations of our study which deserve mention. The generalizability of PatientsLikeMe patient population may not reflect the general population of people with MS as users of health-based internet sites are more likely to be female, younger, and more educated than those sampled from a clinic [48]. Results from interviews are not representative of all people with MS on PatientsLikeMe, nor those who completed the survey, as they were selected to include only those who experienced difficulty obtaining a DMT. Patient-reported explanations for DMT access difficulties are subject to errors in recall and errors in the reconstruction of events, especially among participants who reported MS relapse during the access issue. Finally, as the interview sample size was likely not sufficient to achieve concept saturation, resultant themes should be interpreted with caution.

Conclusions

This study stresses the need for future research to incorporate the patient perspective to better understand barriers to MS treatment access. The evaluation of the long-term impact of DMT access barriers on patient and disease outcomes are needed. Formulary decision makers must consider the patient experience when making DMT coverage decisions. Clinicians should be aware of how patients experience DMT access difficulties and help deliver solutions to them when feasible. The MS patient experience with DMT access will continue to evolve with ongoing policy and payer landscape changes. Hence, frequent feedback from people with MS and stakeholders will be of paramount importance to ensure access to DMTs and to measure the associated impact on outcomes.

Acknowledgments

We wish to acknowledge the following individuals for contributions to this manuscript: Nicole Buechler for assistance with survey administration and analysis, Jamie Granskie for assistance with literature review, Marissa Chiauzzi and Lauren Kaplan for assistance scheduling and transcribing interviews, and Jill M. Serrano, PhD, for assistance with the preparation of this manuscript. The PDDS was provided for use by the NARCOMS Registry. NARCOMS is supported in-part by the Consortium of Multiple Sclerosis Centers (CMSC) and the CMSC Foundation. Novartis Pharmaceuticals Corporation provided funding for this study and reviewed this manuscript.

Authors' Contributions

All authors contributed to the development of the publication and maintained control over the final content. Study concept and design were contributed by JJK and BJK with support from KFS, SV, and DM. BJK and KFS led data collection with assistance from DM, and data analysis and interpretation were performed by BJK and KFS with assistance from JJK, SV, and DM.

Conflicts of Interest

KFS is an employee of PatientsLikeMe and holds stock options in the company. BJK was an employee of PatientsLikeMe and held stock options in the company at the time of the research and manuscript writing. The PatientsLikeMe research team has received research funding (including conference support and consulting fees) from Abbvie, Accorda, Actelion, Alexion, Amgen, AstraZeneca, Avanir, Biogen, Boehringer Ingelheim, Celgene, EMD, Genentech, Genzyme, Janssen, Johnson & Johnson, Merck, Neuraltus, Novartis, Otsuka, Sanofi, Takeda, and UCB. The PatientsLikeMe R&D team has received research grant funding from Kaiser Permanente, the Robert Wood Johnson Foundation, Sage Bionetworks, The AKU Society, and the University of Maryland. DM is a patient member of PatientsLikeMe. JJK was an employee of Novartis Pharmaceuticals Corporation with salary and stocks at the time of the research and manuscript writing. SV was an employee of Thomas Jefferson University at the time of the research and manuscript writing. KJ is an employee of Novartis Pharmaceuticals Corporation.

Multimedia Appendix 1

Survey questions.

[PDF File (Adobe PDF File), 138KB - jmir_v20i10e11168_app1.pdf]

Multimedia Appendix 2

Interview guide.

[PDF File (Adobe PDF File), 72KB - jmir_v20i10e11168_app2.pdf]

References

1. Dilokthornsakul P, Valuck RJ, Nair KV, Corboy JR, Allen RR, Campbell JD. Multiple sclerosis prevalence in the United States commercially insured population. Neurology 2016 Mar 15;86(11):1014-1021 [FREE Full text] [doi: 10.1212/WNL.0000000000002469] [Medline: 26888980]
2. National Multiple Sclerosis Society. 2018. Who Gets MS? (Epidemiology) URL: https://www.nationalmssociety.org/What-is-MS/Who-Gets-MS [accessed 2018-05-25] [WebCite Cache ID 6zgE3CfwL]
3. Adelman G, Rane SG, Villa KF. The cost burden of multiple sclerosis in the United States: a systematic review of the literature. J Med Econ 2013;16(5):639-647. [doi: 10.3111/13696998.2013.778268] [Medline: 23425293]
4. Bøe LHM, Telstad W, Grytten N, Kyte L, Aarseth J, Myhr K, et al. Employment among patients with multiple sclerosis-a population study. PLoS One 2014;9(7):e103317 [FREE Full text] [doi: 10.1371/journal.pone.0103317] [Medline: 25054972]
5. Maroney M, Hunter SF. Implications for multiple sclerosis in the era of the Affordable Care Act: a clinical overview. Am J Manag Care 2014 Dec;20(11 Suppl):S220-S227 [FREE Full text] [Medline: 25734887]
6. Cadden M, Arnett P. Factors Associated with Employment Status in Individuals with Multiple Sclerosis. Int J MS Care 2015;17(6):284-291 [FREE Full text] [doi: 10.7224/1537-2073.2014-057] [Medline: 26664334]
7. Gleason PP, Alexander GC, Starner CI, Ritter ST, Van Houten HK, Gunderson BW, et al. Health plan utilization and costs of specialty drugs within 4 chronic conditions. J Manag Care Pharm 2013 Sep;19(7):542-548 [FREE Full text] [doi: 10.18553/jmcp.2013.19.7.542] [Medline: 23964615]
8. Upshur K. The MS Center of Saint Louis: a practice perspective of the impacts of high cost sharing. Am J Manag Care 2016 Mar;22(4 Suppl):s95-s98 [FREE Full text] [Medline: 27270160]
9. National Multiple Sclerosis Society. 2018. Disease Modification URL: https://www.nationalmssociety.org/For-Professionals/Clinical-Care/Managing-MS/Disease-Modification [accessed 2018-05-25] [WebCite Cache ID 6zgDF7Hzs]
10. Hartung DM, Bourdette DN, Ahmed SM, Whitham RH. The cost of multiple sclerosis drugs in the US and the pharmaceutical industry: Too big to fail? Neurology 2015 May 26;84(21):2185-2192 [FREE Full text] [doi: 10.1212/WNL.0000000000001608] [Medline: 25911108]
11. Owens GM, Olvey EL, Skrepnek GH, Pill MW. Perspectives for managed care organizations on the burden of multiple sclerosis and the cost-benefits of disease-modifying therapies. J Manag Care Pharm 2013;19(1 Suppl A):S41-S53 [FREE Full text] [doi: 10.18553/jmcp.2013.19.s1.S41] [Medline: 23383732]

12. Lotvin AM, Shrank WH, Singh SC, Falit BP, Brennan TA. Specialty medications: traditional and novel tools can address rising spending on these costly drugs. Health Aff (Millwood) 2014 Oct;33(10):1736-1744. [doi: 10.1377/hlthaff.2014.0511] [Medline: 25288417]
13. Happe LE, Clark D, Holliday E, Young T. A systematic literature review assessing the directional impact of managed care formulary restrictions on medication adherence, clinical outcomes, economic outcomes, and health care resource utilization. J Manag Care Spec Pharm 2014 Jul;20(7):677-684 [FREE Full text] [doi: 10.18553/jmcp.2014.20.7.677] [Medline: 24967521]
14. Goldman DP, Joyce GF, Lawless G, Crown WH, Willey V. Benefit design and specialty drug use. Health Aff (Millwood) 2006;25(5):1319-1331. [doi: 10.1377/hlthaff.25.5.1319] [Medline: 16966729]
15. Romley JA, Sanchez Y, Penrod JR, Goldman DP. Survey results show that adults are willing to pay higher insurance premiums for generous coverage of specialty drugs. Health Aff (Millwood) 2012 Apr;31(4):683-690. [doi: 10.1377/hlthaff.2011.1301] [Medline: 22492884]
16. Menzin J, Caon C, Nichols C, White LA, Friedman M, Pill MW. Narrative review of the literature on adherence to disease-modifying therapies among patients with multiple sclerosis. J Manag Care Pharm 2013;19(1 Suppl A):S24-S40 [FREE Full text] [doi: 10.18553/jmcp.2013.19.s1.S24] [Medline: 23383731]
17. Lizán L, Comellas M, Paz S, Poveda JL, Meletiche DM, Polanco C. Treatment adherence and other patient-reported outcomes as cost determinants in multiple sclerosis: a review of the literature. Patient Prefer Adherence 2014;8:1653-1664 [FREE Full text] [doi: 10.2147/PPA.S67253] [Medline: 25525341]
18. Ma VY, Chan L, Carruthers KJ. Incidence, prevalence, costs, and impact on disability of common conditions requiring rehabilitation in the United States: stroke, spinal cord injury, traumatic brain injury, multiple sclerosis, osteoarthritis, rheumatoid arthritis, limb loss, and back pain. Arch Phys Med Rehabil 2014 May;95(5):986-995.e1 [FREE Full text] [doi: 10.1016/j.apmr.2013.10.032] [Medline: 24462839]
19. Wang G, Marrie RA, Salter AR, Fox R, Cofield SS, Tyry T, et al. Health insurance affects the use of disease-modifying therapy in multiple sclerosis. Neurology 2016 Jul 26;87(4):365-374 [FREE Full text] [doi: 10.1212/WNL.0000000000002887] [Medline: 27358338]
20. Howard DH. Drug companies' patient-assistance programs--helping patients or profits? N Engl J Med 2014 Jul 10;371(2):97-99. [doi: 10.1056/NEJMp1401658] [Medline: 25006717]
21. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
22. Wallace LS, Rogers ES, Roskos SE, Holiday DB, Weiss BD. Brief report: screening items to identify patients with limited health literacy skills. J Gen Intern Med 2006 Aug;21(8):874-877. [doi: 10.1111/j.1525-1497.2006.00532.x] [Medline: 16881950]
23. Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. Fam Med 2004 Sep;36(8):588-594 [FREE Full text] [Medline: 15343421]
24. Hohol MJ, Orav EJ, Weiner HL. Disease steps in multiple sclerosis: a longitudinal study comparing disease steps and EDSS to evaluate disease progression. Mult Scler 1999 Oct;5(5):349-354. [doi: 10.1177/135245859900500508] [Medline: 10516779]
25. Hohol MJ, Orav EJ, Weiner HL. Disease steps in multiple sclerosis: a simple approach to evaluate disease progression. Neurology 1995 Feb;45(2):251-255. [Medline: 7854521]
26. Marrie RA, Goldman M. Validity of performance scales for disability assessment in multiple sclerosis. Mult Scler 2007 Nov;13(9):1176-1182. [doi: 10.1177/1352458507078388] [Medline: 17623733]
27. Learmonth YC, Motl RW, Sandroff BM, Pula JH, Cadavid D. Validation of patient determined disease steps (PDDS) scale scores in persons with multiple sclerosis. BMC Neurol 2013 Apr 25;13:37 [FREE Full text] [doi: 10.1186/1471-2377-13-37] [Medline: 23617555]
28. 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]
29. Muhr T, Friese S. User's Manual for ATLAS ti 5.0. In: ATLAS.ti Sci Softw Dev GmbH. Berlin: Scientific Software Development; 2004.
30. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357 [FREE Full text] [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
31. Boeije H. A purposeful approach to the constant comparative method in the analysis of qualitative interviews. Qual Quant 2002;36(4):391-409. [doi: 10.2307/798843]
32. Glaser B. The constant comparative method of qualitative analysis. Soc Probl 1965;12(4):436-445. [doi: 10.2307/798843]
33. Scott J. Social Network Analysis. 2nd ed. Thousand Oaks, CA: Sage; 2000.
34. Knoke D, Yang S. Social Network Analysis. Los Angeles: Sage; 2008.
35. Wicks P, Vaughan T, Heywood J. Subjects no more: what happens when trial participants realize they hold the power? BMJ 2014 Jan 28;348:g368 [FREE Full text] [Medline: 24472779]

36. Eaddy MT, Cook CL, O'Day K, Burch SP, Cantrell CR. How patient cost-sharing trends affect adherence and outcomes: a literature review. P T 2012 Jan;37(1):45-55 [FREE Full text] [Medline: 22346336]
37. Tan H, Cai Q, Agarwal S, Stephenson JJ, Kamat S. Impact of adherence to disease-modifying therapies on clinical and economic outcomes among patients with multiple sclerosis. Adv Ther 2011 Jan;28(1):51-61. [doi: 10.1007/s12325-010-0093-7] [Medline: 21153000]
38. Ivanova JI, Bergman RE, Birnbaum HG, Phillips AL, Stewart M, Meletiche DM. Impact of medication adherence to disease-modifying drugs on severe relapse, and direct and indirect costs among employees with multiple sclerosis in the US. J Med Econ 2012;15(3):601-609. [doi: 10.3111/13696998.2012.667027] [Medline: 22376190]
39. Dor A, Lage MJ, Tarrants ML, Castelli-Haley J. Cost sharing, benefit design, and adherence: the case of multiple sclerosis. Adv Health Econ Health Serv Res 2010;22:175-193. [Medline: 20575233]
40. Iezzoni LI, Ngo LH, Kinkel RP. Working-age persons with multiple sclerosis and access to disease-modifying medications. Mult Scler 2008 Jan;14(1):112-122. [doi: 10.1177/1352458507080466] [Medline: 17893109]
41. Naci H, Fleurence R, Birt J, Duhig A. Economic burden of multiple sclerosis: a systematic review of the literature. Pharmacoeconomics 2010;28(5):363-379. [doi: 10.2165/11532230-000000000-00000] [Medline: 20402540]
42. Minden SL, Frankel D, Hadden L, Perloffp J, Srinath KP, Hoaglin DC. The Sonya Slifka Longitudinal Multiple Sclerosis Study: methods and sample characteristics. Mult Scler 2006 Feb;12(1):24-38. [doi: 10.1191/135248506ms1262oa] [Medline: 16459717]
43. van der Hiele K, van Gorp D, Ruimschotel R, Kamminga N, Visser L, Middelkoop H. Work Participation and Executive Abilities in Patients with Relapsing-Remitting Multiple Sclerosis. PLoS One 2015;10(6):e0129228 [FREE Full text] [doi: 10.1371/journal.pone.0129228] [Medline: 26083386]
44. Bourdette D, Patti F. US health insurance is an obstacle to disease-modifying treatments in MS. Neurology 2016 Jul 26;87(4):346-347. [doi: 10.1212/WNL.0000000000002904] [Medline: 27358336]
45. National Multiple Sclerosis Society. Transparency of Cost and Coverage. 2018. URL: https://www.nationalmssociety.org/Get-Involved/Advocate-for-Change/Current-Advocacy-Issues/Transparency-of-Coverage-and-Cost [accessed 2018-05-25] [WebCite Cache ID 6zgCIHAKq]
46. Kocher R, Emanuel EJ, DeParle NM. The Affordable Care Act and the future of clinical medicine: the opportunities and challenges. Ann Intern Med 2010 Oct 19;153(8):536-539. [doi: 10.7326/0003-4819-153-8-201010190-00274] [Medline: 20733178]
47. Jones JL, Scheidt DJ, Kaushal RS, Carroll CA. Assessing the role of patient support services on adherence rates in patients using glatiramer acetate for relapsing-remitting multiple sclerosis. J Med Econ 2013;16(2):213-220. [doi: 10.3111/13696998.2012.744316] [Medline: 23098539]
48. Bove R, Secor E, Healy BC, Musallam A, Vaughan T, Glanz BI, et al. Evaluation of an online platform for multiple sclerosis research: patient description, validation of severity scale, and exploration of BMI effects on disease course. PLoS One 2013;8(3):e59707 [FREE Full text] [doi: 10.1371/journal.pone.0059707] [Medline: 23527256]

Abbreviations

DMT: disease-modifying therapy
MRI: magnetic resonance imaging
MS: multiple sclerosis
PDDS: Patient-Determined Disease Step
RRMS: relapsing-remitting multiple sclerosis

Edited by G Eysenbach; submitted 29.05.18; peer-reviewed by L Lavorgna, I Alguacil; comments to author 11.07.18; revised version received 18.07.18; accepted 18.07.18; published 30.10.18.

Please cite as:
Simacek KF, Ko JJ, Moreton D, Varga S, Johnson K, Katic BJ
The Impact of Disease-Modifying Therapy Access Barriers on People With Multiple Sclerosis: Mixed-Methods Study
J Med Internet Res 2018;20(10):e11168
URL: http://www.jmir.org/2018/10/e11168/
doi:10.2196/11168
PMID:30377144

Original Paper

Discourse on Exposure to Pornography Content Online Between Arab Adolescents and Parents: Qualitative Study on its Impact on Sexual Education and Behavior

Anat Gesser-Edelsburg[1], PhD; Munawar Abed Elhadi Arabia[2], BA

[1]The Health and Risk Communication Research Center, School of Public Health, University of Haifa, Haifa, Israel

[2]School of Public Health, University of Haifa, Haifa, Israel

Corresponding Author:
Anat Gesser-Edelsburg, PhD
The Health and Risk Communication Research Center
School of Public Health
University of Haifa
199 Aba Khoushy Avenue, Mount Carmel
Haifa, 3498838
Israel
Phone: 972 544243530
Fax: 972 36322143
Email: ageser@univ.haifa.ac.il

Abstract

Background: The internet revolution of the 21st century has made sexual content available and accessible on a scale that has never existed before. Many studies have indicated that the use of pornography was associated with more permissive sexual attitudes and tended to be linked with stronger gender-stereotypical sexual beliefs. It also seemed to be associated with other risky behaviors and sexual promiscuity. Pornography exposure in conservative societies leads to conflicts with religious and cultural taboos.

Objective: The aim of this study was to characterize the barriers and difficulties that prevent sexual discourse in the Arab society and enable pornography viewing according to the perceptions of adolescents and mothers.

Methods: This study involved qualitative research methods and in-depth interviews with 40 participants. This study included 20 Arab adolescents, sampled by 2 age groups (14-16 years and 16-18 years), and 20 mothers of adolescents from both sexes.

Results: The findings indicate that mothers "turn a blind eye" to porn viewing and sexual activity by boys; however, they show a sweeping prohibition and denial of such behavior by girls. Boys reported viewing porn routinely, whereas girls denied doing so, but admitted that their female friends watched porn. The study also found that boys experienced guilt during and after viewing porn as a result of the clash between modernity and traditional values. The mothers and adolescents emphasized the need for an open sexual discourse to reduce violent behaviors such as Web-based sexual harassment, including sending videos and pictures of naked girls, often accompanied by threats and blackmail.

Conclusions: It is necessary to find a way to encourage a significant sexual discourse to prevent the violent consequences of its absence in the Arab society. A controlled, transparent, and critical sexual discourse could help youth make more informed decisions concerning the search for sexual content, porn viewing, and sexual behavior.

(J Med Internet Res 2018;20(10):e11667) doi:10.2196/11667

KEYWORDS

internet; pornography; porn viewing; sexuality; taboo; Israeli Arab adolescents; discourse; sexual education and behavior; sexual double standard

Introduction

Sexual Discourse in the Arab Community: Between Tradition and Modernization

In the Arab Community, Sexual Discourse is Considered Taboo

Among Islamic religious individuals, explicit sexual discourse is not encouraged. The reason is that according to the religion and rules of the Quran, only married people are permitted to engage in sexual relations, and therefore, sexual discourse with adolescents is considered to potentially encourage premarital sex [1]. However, as Roudi-Fahimi [2] indicated in their systematic review, despite the religious prohibition, there is in fact sexual contact between young people. Therefore, awareness has grown in some of the Arab countries that the population should be given access to information about contraceptives and about sexually transmitted diseases (STDs) resulting from multiple partners [2]. However, in contrast with the recognition of the sexual activity of youth by health authorities in Arab countries, there is still a social taboo and denial by parents that Arab youth have premarital sex, and therefore, sexual discourse in the public and domestic spheres is nonexistent [3].

In reality today, Arab youth experience sexual contact, porn viewing, and premarital sex [3-6]. The main reasons are that in the last decade, Arab communities have undergone immense transformations, largely related to the infiltration of Western elements into the society, technological means that narrowed the gap between Arab and Western society, and global social and economic changes [7,8]. The process of modernization where women education increases, fertility decreases, and marital timing is postponed occurs mainly in Arab communities living in advanced democratic countries. Furthermore, Arab youth are more exposed than the older generation to the effects of the technologies that arrived with the new media revolution. Arab youth are currently living in a dual reality where on the other hand, they are attracted and yearning for the liberal technological Western culture, while at the same time, they want to maintain their allegiance to their ancestors' traditional isolationist culture [8-10].

Studies have shown that the nature of sexual discourse, its culture, and the way it is managed have a predictive correlation with the actual behavior of adolescents and adults [11,12]. The absence of overt sexual discourse also leads to ignorance [13], fears, and anxiety in Arab adolescent boys and girls. For instance, many Arab girls reported that the appearance of signs of menstruation came to them as a complete surprise [13,14].

Sexual Double Standard and the Status of Women in Arab Society

The sexual double standard is the widespread belief that sexual behaviors are judged differently depending on the gender of the sexual actor [15]. The sexual double standard is policed and controlled through "the-male-in-the-head" [16]. This concept refers to male power under heterosexuality, which leads to the unequal relationship between femininity and masculinity, and relates to the control of both female and male sexuality. A characteristic of the-male-in-the-head is the silencing of sexual female voices and the simultaneous noise of male-dominated conversations in this area. Boys and men are thought to receive praise and positive attributions from others for nonmarital sexual contacts, whereas girls and women are believed to be derogated and stigmatized for similar behaviors. In other words, men are rewarded for sexual activity, whereas women are derogated for the same activity [17,18]. The sexual double standard is related to standard gender stereotypes: sex and desire are not feminine, whereas they are expected from men. Heterosexuality is constructed under a male gaze [19] so that men are in the position of power and they have access to discourses of sex and desire, whereas women's desire is silenced. Women are supposed to hide their desire and make it invisible [19], whereas heterosexual men can express it openly. Furthermore, sex is seen as a greater risk for women because they can become pregnant, and although men can easily walk away from this situation, women have to shoulder the responsibility [18].

The sexual double standard intensifies in patriarchal societies such as the Arab society. In the Arab society, the woman is considered property of the man. Not only is her status unequal, but the desires of the man dictate her behavior. Women's expression of sexual wishes or desires that contradict those of the man are often considered an offense against the man's honor and the family honor [20].

Thus, men, given their superior position and their perception of women as their property, often rape women. It is important to emphasize that the very concept of rape does not exist in many Arab countries and that the act that is punishable is extramarital sex (a man is definitely allowed to rape his wife) [21,22]. According to the laws in these countries, 4 witnesses are often required to make a charge of rape. In the absence of 4 witnesses, the most important evidence to support a charge of rape or, alternately, extramarital sex is the woman's pregnancy as a result of the rape. The woman is accused and punished for being raped, whereas the man is not accused at all. The situation clearly discriminates against women and places a double punishment on the victim [23]. In some Arab countries, to "ease" the woman's plight, she is not punished but rather ordered to raise the child without a father (of course abortion is not permitted, even in cases of rape) [24]. Another issue associated with double standards toward women is murder on the background of so-called family honor, which is also familiar in the Arab society in Israel. Although men are not judged for having sex before or outside of marriage, women are judged by the society and even murdered for what is defined as "improper sexual behavior" [25].

The Internet as a Source of Sexual Information and Consumption

Internet-enabled devices have enabled people of all ages to consume sexual information at an availability and speed that have impacted and changed the sexual habits and knowledge of adolescents [26,27].

The internet is considered a more highly sexualized environment than other media [28], and research has shown that the number of youth who intentionally or accidentally encounter pornographic material online has risen significantly [29,30].

The internet occupies a prominent and prioritized place in the lives of many youth [29,31,32]. For example, a survey by the World Internet Report of people aged 12 to 14 years from 13 different countries found that 100% of British, 98% of Israeli, 96% of Czech, and 95% of Canadian youth reported using the internet regularly [33].

The internet can serve youth as a source for the acquisition of skills, the development of high literacy, and entertainment [34]. However, at the same time, for users with certain sociodemographic characteristics, it can be a source of risk behaviors, such as pornography viewing and addiction [35,36].

Pornography and Youth

The legal status of pornography in the world varies widely from one country to another [37], but attempts to restrict access to online pornographic contents in different countries have usually failed due to the ease of access [38]. Systematic studies and reviews have indicated that youth view porn from 10 years to 18 years, although prevalence rates between studies varied greatly [39].

Young adults can also be exposed to pornographic contents "by chance" when they do not intend to [40-42]. Exposure to sexually explicit material in adolescence has a unique impact because in adolescence, youth feel high uncertainty about their identity and sexual boundaries [43]. Furthermore, exposure to porn from a young age impacts the way youth think about sexuality as well as their actual sexual behavior. According to a large survey of American college students, 51% of males and 32% of females admitted to viewing pornography for the first time before they were 13 years old [44]. For teenagers exposed to pornography within a family setting, pornography causes stress and increases the risk for developing negative attitudes about the nature and purpose of human sexuality.

For adolescents who view pornography, their attitudes toward their own and others' sexuality change, and their sexual expectations and behavior are shaped accordingly [43-45]. A study of 2343 adolescents found that sexually explicit internet material significantly increased their uncertainties about sexuality [43].

Age groups of 14-16 years and 16-18 years are sensitive ages for porn viewing because from age 14 years, adolescents face growing social pressure from their peer group to have romantic partners [46,47]. Relations with partners at these ages are influenced by what they viewed and learned from porn.

Due to the pervasiveness of online porn in the cultural and social environment of adolescents, a systematic survey was held in 2016 [48], which found (despite the differences in the methodologies of the different studies) that pornography use was associated with more permissive sexual attitudes and tended to be linked with stronger gender-stereotypical sexual beliefs. It also appeared to be related to the occurrence of sexual intercourse, greater experience with casual sexual behavior, and more sexual aggression, both in terms of perpetration and victimization.

Porn viewing often leads to adolescents' lower levels of sexual self-esteem [49], more liberal sexual positions, and a higher belief that peers are sexually active, raising the likelihood of younger sexual initiation [26].

Adolescents who are exposed to sexual behaviors outside of cultural norms may develop a distorted perception of sex as unrelated to love and intimacy and a desire for sexual engagement without emotional commitment [50]. The combination of peer pressure, porn viewing, and patriarchal values leads to risky behavior [51].

Studies have indicated that mostly boys but also girls tend to engage in more "sexting" (exchanging sexual text messages) when they view porn. Sexting by youth often leads to sexual contempt and online sexual violence. Studies indicate that when sexting is accompanied by alcohol drinking, it leads youth to a loss of control and potential sexual violence [52,53]. Moreover, adolescents who are exposed to pornography may develop positions that support the "rape myth," which ascribes responsibility for sexual assault to the female victim [26,54].

There are few studies on the viewing habits of sexual contents and use of pornography in Arab countries among young adults in general and adolescents in particular. Studies that have examined the subject have found that the internet exposes Arab youth to contents that contradict religious and cultural taboos. The studies found that because of the proscription and supervision in Arab countries, youth acquire information and view pornography clandestinely [55].

Arab youth living in conservative communities watch porn secretly not only out of the psychological fear of the reactions of their parents and other authority figures in their lives [43] but also because of the religious proscription that does not exist for secular youth living in liberal societies [56].

It was found that because Arab adolescents live in a conservative world with a culture of silencing, their level of emotional readiness and tools to filter sexual contents are much lower than that of Western youth [13,57]. For instance, in a study of young adult students in Lebanon, it was found that a considerable number of them used the internet to view pornography and gamble [58].

Many studies are conducted around the world on pornography consumption. However, as stated in the systematic survey by Owens et al [59], it is necessary to continue studying this phenomenon by encouraging global studies. Few studies have been conducted on Arab youth concerning sexual discourse on the social networks and the use of technological means. Most of the studies in the literature are quantitative studies that indicate the frequency of porn viewing and/or the attitudes and perceptions of youth about specific issues on which they were asked in closed questionnaires. There are very few qualitative research studies, namely, in-depth "face-to-face" interviews that take a deep look at the features of the sexual discourse between Arab youth and their parents as well as the gaps and conflicts that arise from them.

In light of the sparsity of data specifically concerning this population, this study will contribute to understanding the impact of social, cultural, and religious patterns on the perceptions of online sexual discourse of Arab youth and mothers in Israel. Moreover, the study can provide a basis for the formulation of recommendations that put an emphasis on risk communication on the internet with the purpose of promoting a policy of effective and accommodated sexual discourse for the needs of Arab youth and their parents.

Objectives

The objective of this study was to characterize the barriers and difficulties that prevent sexual discourse in the Arab society and enable pornography viewing according to the perceptions of adolescents and mothers.

Methods

Study Design and Analysis

This study employs qualitative research, which allows the in-depth observation of a phenomenon through the prism of the study participants. The purpose of the phenomenological qualitative research method is to understand the studied phenomenon by analyzing the experiences of a certain population, with an emphasis on selecting an informative group that authentically represents it [60].

Research Population

A total of 40 respondents were interviewed for this study. This study comprised 20 Arab adolescents (Table 1) in 2 age groups, who, according to the literature, are in different developmental stages: 14-16 years and 16-18 years [61]. In addition, 20 mothers (Table 2) of adolescents of both sexes were interviewed. Only mothers were chosen and not fathers on the assumption that men in the Arab society would refuse to have a conversation on sexuality in general, and especially the fathers of daughters.

Recruitment and Interview Process

An application was submitted to the Faculty of Social Welfare and Health Sciences Ethics Committee for research with human subjects at Haifa University, and full ethical approval (no.439/17) was granted. Participants were recruited through purpose sampling of Arab schools in Nazareth, Kafr Sullam, Reina, Kafr Nin, and Ein Mahel. These schools were chosen to achieve a heterogenic profile of the youth population. The researchers approached different schools in Nazareth and its surroundings to sample youth from different ethnicities—Muslims and Christians. It is important to note that the youth in Nazareth live in a mixed urban environment including Jews. This environment is fundamentally different from the isolation of the purely Arab population such as in the villages of Kafr Sullam, Reina, Kafr Nin, and Ein Mahel.

The mothers of boys and girls were approached through class WhatsApp groups. The approach laid out the research goal and provided the contact information of 1 of the researchers and an invitation to contact her. The researcher asked the mothers for permission to interview their children. Following the mothers' approval, the researcher contacted the adolescents and asked for their consent to participate in the study. In addition, mothers were approached separately. It should be noted that it was decided not to interview the adolescents whose mothers agreed to be interviewed to allow interviewees to speak freely. The interviews with the adolescents were conducted wherever the interviewees felt comfortable, usually in their homes or in parks.

Table 1. Adolescent interviewees: sociodemographic data.

Sociodemographic characteristics	Statistics
Gender, n (%)	
Men	11 (55)
Female	9 (45)
Age (years), mean (SD); range	16.1 (1.1); 14.0-18.0
School level (grade), n (%)	
8	2 (10)
9	3 (15)
10	2 (10)
11	10 (50)
12	2 (15)
Religion, n (%)	
Muslim	19 (95)
Christian	1 (5)
Place of residence, n (%)	
Nazareth	15 (75)
Kafr Sullam	3 (15)
Reina	1 (5)
Kafr Nin	1 (5)

Table 2. Sociodemographic data of mother interviewees[a].

Sociodemographic characteristics	Statistics
Age (years), mean (SD); range	41.1 (3.7); 35.0-47.0
Education (years), mean (SD); range	13.6 (2.3), 11.0-20.0
Employment, n (%)	
Income tax	2 (10)
Teacher	3 (15)
Housewife	8 (40)
Nurse	2 (10)
Cashier	2 (10)
Clerk	1 (5)
Secretary	1 (5)
Doctor	1 (5)
Place of residence, n (%)	
Ein Mahel	1 (5)
Nazareth	17 (85)
Daburiyya	1 (5)
Yafia	1 (5)
Level of religiousness, n (%)	
Secular	3 (15)
Traditional	10 (50)
Religious	7 (35)
Income level, n (%)	
Low	3 (15)
Medium	12 (60)
High	5 (25)
Number of children, mean (SD); range	3.6 (1.6); 1-6

[a] All mother interviewees were born in Israel and they are Muslim by religion.

The interviews with the mothers were conducted in their homes. The interviews lasted between 45 min and 1 hour and were conducted by 1 of the researchers who was trained to conduct qualitative interviews. The interviews were recorded and transcribed.

Research Tools

The choice of personal interviews rather than focus groups was made to give the interviewees the confidence to speak freely about a sensitive subject. Semistructured protocols were prepared for the interviews, adjusted to the research subpopulations. The interviews were held in Arabic, the participants' mother tongue. Moreover, 2 protocols were designed for this study: for the adolescents and the mothers. The protocols for the Arab adolescents included questions on the perception of sexual discourse with peers and parents, information searching about sex and sexuality, and pornography viewing. The protocols for the interviews with the mothers included questions about their relationships with their adolescent children, sexual discourse at home, sources of information about their children's sexuality, and sex education.

Data Analysis

The findings were analyzed by the content analysis approach [62] using the following process: in the first stage, themes were analyzed and coded for each population, adolescents and mothers, separately, while identifying main themes and subthemes. In the second stage, themes that arose among the 3 research groups—adolescents aged 14-16 years, adolescents aged 16-18 years, and mothers—were analyzed and coded. In the third stage, each subgroup was integrated separately; all of the interviews of the adolescents in each age group and of the mothers were integrated separately. In the final stage, integrated super-categories were constructed for all of the research populations.

Validity and Reliability

Interviews were recorded, transcribed, and logged in a field diary. This enabled examination of the reliability of the data received from the participants and control of the analysis of the findings by the researchers [63].

The field diary included notes of the time and the place of the interview, the dynamics during the meeting, resistance by interviewees to questions in the interview, and nonverbal reactions (such as body gestures or facial expressions) that cannot be surmised from the transcript of the interview. Given the sensitivity of the subject of sexuality both for the youth and the mothers, the researchers' documentation and reflection on the process was a tool for correcting and improving the discourse with the interviewees as well as providing a holistic and deeper picture of the data.

The interview protocols were designed in Hebrew and translated into Arabic, the mother tongue of the research population, and then translated back from Arabic into Hebrew to check the wording. The interviews were transcribed in Arabic by 1 of the researchers who is fluent in both Arabic and Hebrew. Likewise, a number of stages of data collection and analysis were performed: a pilot to test protocols with 2 mothers and 2 adolescents, joint meetings of the researchers during the data collection process, reading of transcripts by the 2 researchers separately, and decision of categories and subthemes through an agreement between the researchers. Moreover, study participants represented different subpopulations (adolescents by age groups and mothers), which can strengthen the credibility and validity of the findings in relation to the studied phenomenon [62].

Results

Main Findings

The main findings that arose from the interviews with the youth and the mothers indicate 4 central themes. The first theme is the absence of sexual discourse between adolescents and their parents. The technological internet revolution has led to the availability and accessibility of sexual contents but did not advance the discourse between youth and their parents, and sexual discourse is still a social taboo. The second theme included barriers that prevent sexual discourse: normative, religious, cultural, and psychological (see details below). The third theme is that the internet realm presents a unique conflict for Arab youth from a conservative society between attraction to pornography and traditional norms. The fourth theme is the consequences of porn viewing—sexual aggression.

Absence of Sexual Discourse Between Adolescents and Their Parents

All the adolescents (n=20), without exception, stressed that sex and sexuality are taboo and there is no sexual discourse between them and their parents. For instance, 1 of the boys said:

> *In our society parents do not talk about sex. They perceive the subject as sensitive and forbidden, and therefore we as teenagers look for another way to understand the world of sex...*

Similarly, Arab mothers (n=20) also stressed the fact that the subject of sexuality and sexual discourse is a social taboo and that is one of the reasons for the absence of sexual discourse with their children. For instance, 1 of the mothers said:

> *I don't know any parents who have a sexual discourse with their teenaged children. In our society it is forbidden to talk about it. You leave it until they get married and then they learn everything by themselves... Our society does not talk about such things.*

Most of the mothers in the study (n=18) had instrumental sexual discourse with their daughters, which was limited to physiological development, but they did not discuss physiological changes with their sons. One of the mothers said she explains the physiological changes to her daughters and lets her husband talk to their sons:

> *Yes, we do discuss issues related to adolescence, the changes that occur in your body, I discuss "periods" with my daughters more than with my sons. I don't talk to them, it's hard for me! When it comes to boys, I leave it to their father, even though most of the time he does not show any interest.*

The interviews in the study found that some mothers (n=14) emphasized that the conversation with the boys focused only on STDs to warn and scare them about the consequences of having sex "before marriage." For example, 1 of the mothers of sons said:

> *The most important thing for me is to talk about STDs like AIDS. I keep on scaring him that it is an incurable disease. Whoever gets AIDS has a slow death, is rejected by our society. A person who has that disease is perceived as disgusting, perverted, and having had "forbidden" sex. I use the intimidation mechanism to make sure he doesn't have sex.*

Table 3 presents the barriers raised by interviewees concerning the absence of sexual discourse with their children.

Conflict: Attraction to Pornography Against Traditional Norms

The adolescents said that curiosity and the absence of discourse at home led most of them to seek information on the internet and especially to watch porn. All the boys in the interviews (n=10) reported that they watch pornographic movies. For example, 1 of the boys said:

> *My friends at school go into those sites... Porn sites. They watch everything that has to do with sex. Intercourse and so on. Because they want to get to know that world.*

As for girls, a more complex picture arose from the interviews. On one hand, most of the girls (n=6) denied viewing porn, but on the other hand, all the girls stated that their female friends did. It can be assumed that not all the girls indeed watch porn, but because of embarrassment to admit to it directly, they prefer to state that their female friends do it. In addition, the girls state their attraction and repulsion over dealing with sexuality.

Table 3. Barriers that prevent sexual discourse.

Barrier and subheadings	Selected response
The normative barrier	
Conducting sexual discourse is forbidden	"Our society is very conservative. They see this kind of subject as something wrong."
Issues and dilemmas could not be raised in the public domain and social networks	"I don't know, if the media talks about such subjects I think it is done very superficially because otherwise they would get unbearable criticism."
The religious barrier	
Sex is "forbidden in Islam"	"No, of course not. It is a very sensitive subject. Thank God me and my daughters are religious. We don't think about sex or sexuality."
Effect of religion on the absence of discourse	"Religion is also a main factor. The Islamic religion does not allow us and our parents to talk about sex. We know that it is wrong to view those movies because even viewing is adultery according to Islam. It is called 'visual adultery' but we still do it."
The cultural barrier	
Absence of discourse is passed through generation	"Because we got used to it. Our parents did not talk to me about anything to do with sex. So I don't talk either."
Awareness among adolescents of the history of absence	"Our parents don't talk, they are afraid of such subjects, because they got used to it. Their parents didn't talk either. So they feel that such a conversation is wrong."
Discourse as a psychological barrier	
Shyness and unacceptability of sexual discourse	"Sexual discourse is not acceptable in our society like other subjects. It comes from the culture of shyness we grew up in. We are the parents, mothers and fathers, so how could we transmit or talk about such a sensitive subject with our children if we don't have the right education ourselves? We learned about this from our personal experience."
Embarrassment over touching sexual subjects among adolescents	"There are lots of boys who are embarrassed to ask their parents and there are some who are afraid of punishment."
Giving legitimacy to sexual activity	
Concerns and fears of encouraging unacceptable sexual behaviors	"No, I did not talk about safe sex, like using a condom, because that would encourage them. As a result, I prefer not to talk about safe sex. If I am talking about a condom then he won't listen to me when I talk about sex being forbidden and STDs[a]. If I talk about a condom then he will rely on it and won't mind having sex."
The age barrier	
Children aren't mature enough to talk with them about sex (age group 14-16)	"No, my husband and I do not talk about sex. I think there is time for that. But I saw my son watch sexual movies! Yes, I punished him. Because he is still too young for such things. They're pretty young, 9, 14. I think it doesn't make sense to talk about sex. I'm sure they can't process it. Why confuse them?"
Fear of their husbands' reactions to having sexual discourse at home	
Said sexual conversations at home are not at all acceptable to husbands	"No way, of course not!! My husband would kill me if he heard such a thing at home. He would start shouting right away, breaking everything he saw, that's why I prefer not to talk about such things. We got used to not talking with the children about things like love and sexuality."

[a] STDs: sexually transmitted diseases.

For example, 1 interviewee said:

> *I always thought that pregnancy occurred the moment a man and a woman kiss. Or when a woman drinks water from the man's glass. They explained to me that my information was wrong. They told me the truth. I didn't like the conversation and as a result I left the conversation/the group.*

The interviews indicate that most of the adolescents expressed an inner conflict between their attraction to viewing porn and traditional values. Most of the boys (n=9) reported feeling guilty because of the conservative education they received from their society and parents. For example, 1 of the boys stressed:

> *On the one hand we know it is forbidden, on the other hand we want it and need it. And you feel guilty every time you watch.*

Similarly, another boy shared that:

> *There is an internal conflict and problem of conscience, because on the one hand the boys want to watch movies and know everything, to experience the experience and feelings, and on the other hand they know it is wrong and forbidden by the religion, our parents don't accept it.*

There were boys (n=7) who reported that they do not feel guilty while watching porn, but they do feel guilty only after they finish watching it:

> *While watching there is no conflict because we are focused on the movie. The inner conflict, the guilt, between knowing that it is forbidden and consuming porn, appears after the movie is over.*

As aforesaid, the girls said that their friends watch but they do not. They mentioned the guilt that goes with watching porn. One interviewee said:

> *I think they feel guilty, because they know this is all against our culture and values. I'm sure the conflict is much worse for the girls, because our society puts an emphasis and is afraid of anything that happens to a girl. You know and probably heard about cases of girls being murdered, therefore girls do it secretly and experience greater conflict.*

The interviews with the mothers indicated that the boys' mothers know that they watch porn, whereas the girls' mothers tended to deny that their girls did so. One mother said there is a difference between what the patriarchal Arab society allows boys and girls:

> *We as mothers are aware that our boys watch porn and talk about what they watched with each other, but we ignore it and move on! But in Arab society that is not the case for girls. We impose all of the housework on them, in addition to schoolwork, so that they don't have time to think about 'sexual desire'. Some prefer to marry them off young to maintain family honor.*

It emerged among the adolescents that despite risky behaviors such as drinking alcohol and viewing porn, premarital sex is still a significant barrier for them. Boys (n=9) noted that they oppose premarital sex because it disrupts the correct order of relations:

> *Of course, I'm against premarital sex because if we do it before marriage, the desire for marriage goes down, and in the end most of the youth will not get married.*

Some of the youth (n=18) and the mothers (n=20) explained that they opposed premarital sex because of the Islamic religion, which proscribes sexual relations without religious sanction before marriage. One mother said:

> *I'm against premarital sex. First of all it's forbidden by our religion. Secondly it's unacceptable in our society. Thirdly I think it violates the trust between the girl and her parents.*

Family honor is also 1 of the main barriers that prevent youth from having premarital sex. One of the boys described it as follows:

> *Our society does not accept it. It is "merciless" and if they find someone who had sex the result is "suicide" or banishing from a certain area.*

Moreover, 1 of the boys said that if a girl had premarital sex she would be portrayed as "used goods":

> *Men are allowed to do everything, even premarital sex. On the other hand, girls are not allowed to have premarital sex because otherwise they are perceived as secondhand.*

Likewise, girls said that if a girl becomes pregnant before marriage, she has no future. For instance:

> *Our parents taught us that a girl who has sex before marriage will never get married. Because no man will accept that.*

As for getting pregnant before marriage, all the mothers, especially mothers of girls (n=17), emphasized the sensitivity of the subject and said such an event could have a heavy price.

The adolescents emphasized that a girl who gets pregnant before marriage does not go to her parents to find a solution to her problem. Some of the boys (n=8) declared that the girl would ask for help from her boyfriend. For instance:

> *I think she would go to whoever she had sex with and they would think together about how to get an abortion. If the boy denies or evades then I think she would go either to her girlfriend or sister. Or she would hide and conceal the pregnancy and abort it without anyone knowing.*

Other adolescents, especially girls (n=9), thought the girl would not go to anyone for help because nobody could help her. For instance:

> *That's a very difficult situation. I don't know if she would talk, nobody could help her, I think she would find a solution by herself.*

However, some of the girls emphasized that despite fear of their parents, they would be the only people who could help the girl:

> *How hard it is also depends on her age. If she were 18 it would be less complicated than if she were 16 or 17. I think she would go to her parents because in such a situation only her parents would be able to help.*

The Consequences of Porn Viewing—Blackmail and Sexual Abuse

Although the mothers of boys turned a blind eye, most of the mothers (n=16) expressed fears and concerns about the movies their sons watched and their consequences for their children's sex education:

> *You need to understand that life isn't like a movie. Both the sex and the way they have sex are presented in a really disgusting way, and as a result they look at sex completely differently than in life. I don't think the movies they watch have fair information. Viewing causes addiction and divorce. I know about a lot of cases when a husband and wife broke up because he asked her to do things like what he watched. This will cause conflict and end with divorce.*

According to adolescents, unchecked exposure to pornographic movies and sexual contents also leads to blackmail, sexual harassment, and sexual abuse. For instance, most of the adolescents (n=18) mentioned online sexual harassment by sending videos and pictures of naked girls. One of the boys said:

> *Sexual abuse is not only rape, today there are many cases of boys and girls threatening and blackmailing each other, such as with pornographic pictures and films. Today there is a phenomenon of girls sending pictures of themselves naked.*

One mother discussed how the absence of sexual discourse leads to sexual exploitation and abuse in the Arab society:

> *Frequently we do not allow the girls to know about sex, and on the other hand we are aware that all of the boys watch and look for sexual information online. I'm talking about pornographic sites. Most of the information they get is wrong. It is reflected by the rate of sexual harassment and cases of rape, which we hear about every day on the news.*

Most of the mothers of girls (n=16) emphasized the vital importance of warning their daughters against online sexual abuse because of the availability and accessibility of apps that help rapidly distribute all of those illicit pictures and videos:

> *I told her we are not allowed to take pictures of ourselves and send to WhatsApp groups because there are a lot of people who take advantage of those pictures and change them.*

Boys (n=9) and girls (n=7) stressed that the absence of sex education at school leads them to look for information from other sources and that sex education in the framework of the school could potentially help youths:

> *It is very important to talk in school because our society does not talk and does not let us talk about sex or sex education. There is no awareness of these sensitive issues. We as teenagers go and look in the wrong places. With sex education you could raise a whole generation with a better outlook on sex.*

Another advantage of sex education that was mentioned by a considerable number of boys (n=10) has to do with reducing sexual abuse and other risky sexual behaviors. For instance:

> *I think it is very important because we as boys and girls don't know where to get this information. Our parents don't talk and they don't talk in school either. There needs to be at least one source to aim us in the right direction. And there is a good chance that lectures about sex education would reduce the cases of sexual abuse, rape and so on.*

Discussion

Principal Findings

In spite of many changes occurring in Arab societies as a result of their connection with the West, the subject of sexuality is still a taboo [64]. The taboo was reflected in this study as well. The adolescents and the mothers in this study mentioned the religious, cultural, and psychological barriers that make it difficult for them to discuss sexuality in the family setting. The discourse of sexuality in adolescence is limited only to certain physiological aspects such as girls getting their periods. The main perception is that sexuality must not be discussed, that premarital sexual contact is forbidden by religion, and that sexual discourse can legitimize premarital sex. The literature indicates that despite the religious and cultural proscriptions, Arab youth do have premarital sex [4,64]. The interviews with adolescents and mothers in this study also found that the actual reality is different from the conservative perception. On the surface, the mothers note that the proscription of sexual activity by adolescents is proscripted, but mothers' warnings to their boys to be careful of sexual relations because of the fear of contracting STDs indicate that they take into account that adolescent boys will have premarital sex. The findings of this study strongly indicate the prevailing patriarchal worldview [22,65].

The findings of this study indicate the phenomenon of "sexual double standard," meaning the widespread belief that sexual behaviors are judged differently depending on the gender of the sexual actor [15]. Boys and men are thought to receive praise and positive attributions from others for nonmarital sexual contacts, whereas girls and women are believed to be derogated and stigmatized for similar behaviors. In other words, men are rewarded for sexual activity, whereas women are derogated for the same activity [17,18]. Similarly, in this study, boys and girls said that the girl is the one who would pay the highest price. It is the girl who would be condemned familially and socially; moreover, her life could be in danger, as a result of compromising family honor. The findings presented in this study are compatible with the research literature that shows that men in the Arab society have much more sexual freedom to act without harming family relations, as opposed to women who must subdue to many dictates to maintain domestic peace and family honor [65]. It is important to note that the patriarchy of the Arab society is also reflected by its reported online pornography consumption [66]. The boys in this study reported that they viewed porn, as opposed to the girls, who denied doing so but admitted to doing so indirectly by reporting that their female friends did.

These findings indicate the internalization of the gender stereotype by the youth, that is, sex and desire are not feminine; however, they are expected from men. Heterosexuality is constructed under a male gaze [19]. Thus, men are in the position of power and they have access to discourses of sex and desire, whereas women's desire is silenced. In addition, the findings of the study indicate the internalization of the sexual double standard by the mothers in this research. As Milhausen and Herold [15] point out, men are not the only ones who internalize double standards—in many cases, women do so as well.

Mothers in this study tended to ignore the fact that their sons viewed porn; however, they denied that their daughters might behave the same way. Supposedly, there is a sweeping prohibition against sexual contact and porn watching both for boys and girls, but mothers' lenient attitude toward the behavior of male adolescents emphasizes the prevailing objectification of adolescent girls. It is precisely the mothers, female adults, who are the ones who internalize the patriarchal view. They maintain that it is mainly women who are under the imperative to avoid becoming "bad" girls who have sexual desire and engage in sex with whomever they please [19]. They maintain that women should be judged more harshly than men for sexual activity and that women should "respect" themselves more [67].

Moreover, some of the mothers in this study reported that they avoid talking with their children because they are afraid of the wrath of the father in the family who would not tolerate such conversation. In addition, what arises from this study is the overt and covert discourse that interacts with other risky behaviors in the Arab society, namely, the sweeping prohibition versus what actually happens. For example, the prohibition in Islam on drinking alcohol versus the secret drinking by Muslim youths while parents turn a blind eye [68].

The findings of this study also indicate the ambivalence and internal conflict youth feel about viewing porn. Adolescents feel guilty during and after viewing. They say these feelings arise because of the moral conflict between modernity and traditional values. The fierce internal conflict they feel corresponds with the studies that indicate the duality in which Arab adolescents experience the clash between modernization and traditional values [10]. This clash is reinforced by the new media revolution, which made sexually explicit contents accessible in a way that no other media had ever done before. Furthermore, viewing porn influences the way youth discuss sex among themselves and the way they actually behave. Adolescents reported sexual abuse that occurs in their social realm following porn viewing. Unchecked exposure to pornographic movies and sexual contents leads, according to the interviewed adolescents, also to blackmail, sexual harassment, and sexual abuse. These behaviors have also been found in other studies of youth throughout the world [12,59,69] and in the Arab society in particular.

Limitations

The limitations of this study are that it is a qualitative study, and therefore, it cannot represent the whole population. However, only qualitative research could make it possible to conduct an in-depth conversation about sexuality, an issue which is a social taboo. Given the extreme sensitivity of the subject, interviews could not be held with fathers.

Follow-up studies might also manage to include interviews with fathers to shed light on the issues of sexual discourse and porn viewing. It is very important to try to conduct follow-up studies on the way the sexual discourse is conducted and how it influences sexual behavior and domestic violence. Follow-up studies might design a quantitative measure assessing risky behaviors in the different teen groups.

Conclusions

It is evident in light of the studies that this struggle between the conservative and modern cultures, which plays out within the psyches of the adolescents; the absence of sex education; adolescents' need to search for information; and their unchecked exposure to online porn all highlight the need to change the discourse and provide effective tools to deal with this conflictual situation. The conclusion and recommendation that arises from the study is that it is not enough to transmit information and factual data as has been done so far by the school system. It is necessary to find a way to encourage a meaningful conversation to prevent the violent consequences of its absence. Introducing a sexual discourse and managing it in a controlled, transparent, and critical manner could help youth make more informed decisions concerning the search for sexual contents, porn viewing, and sexual behavior.

Acknowledgments

The authors would like to thank the study participants and the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper.

Conflicts of Interest

None declared.

References

1. Gańczak M, Barss P, Alfaresi F, Almazrouei S, Muraddad A, Al-Maskari F. Break the silence: HIV/AIDS knowledge, attitudes, and educational needs among Arab university students in United Arab Emirates. J Adolesc Health 2007 Jun;40(6):572.e1-572.e8. [doi: 10.1016/j.jadohealth.2007.01.011] [Medline: 17531765]
2. Roudi-Fahimi F. Washington, DC; 2003 Feb. Women's reproductive health in the Middle East and North Africa URL: https://assets.prb.org/pdf/WomensReproHealth_Eng.pdf [accessed 2018-10-01] [WebCite Cache ID 72qNmI0g9]
3. Agha S. Changes in the timing of sexual initiation among young Muslim and Christian women in Nigeria. Arch Sex Behav 2009 Dec;38(6):899-908. [doi: 10.1007/s10508-008-9395-0] [Medline: 18712467]
4. Dialmy A, Uhlmann AJ. Sexuality in contemporary Arab society. Soc Anal 2005;49(2):16-33 [FREE Full text]
5. Foster AM, Wynn L, Rouhana A, Polis C, Trussell J. Reproductive health, the Arab world and the internet: usage patterns of an Arabic-language emergency contraception web site. Contraception 2005 Aug;72(2):130-137. [doi: 10.1016/j.contraception.2005.03.003] [Medline: 16022853]
6. Roudi-Fahimi F, El Feki S. Washington, DC; 2011. Facts of life: youth sexuality and reproductive health in the Middle East and North Africa URL: https://assets.prb.org/pdf11/facts-of-life-youth-in-middle-east.pdf? [accessed 2018-10-01] [WebCite Cache ID 72qObwbVx]

7. Hamade SN. Internet addiction among university students in Kuwait. Domes 2010 Mar 16;18(2):4-16. [doi: 10.1111/j.1949-3606.2009.tb01101.x]
8. Kheirkhah F, Ghabeli Juibary A, Gouran A, Hashemi S. Internet addiction, prevalence and epidemiological features: First study in Iran. Eur Psychiatry 2008 Apr;23(Supplement 2):S309. [doi: 10.1016/j.eurpsy.2008.01.1063]
9. Massad SG, Karam R, Brown R, Glick P, Shaheen M, Linnemayr S, et al. Perceptions of sexual risk behavior among Palestinian youth in the West Bank: a qualitative investigation. BMC Public Health 2014 Nov 24;14:1213 [FREE Full text] [doi: 10.1186/1471-2458-14-1213] [Medline: 25420532]
10. Zeira A, Astor RA, Benbenishty R. Sexual harassment in Jewish and Arab public schools in Israel. Child Abuse Negl 2002 Feb;26(2):149-166. [doi: 10.1016/S0145-2134(01)00314-3]
11. Fix RL, Falligant JM, Alexander AA, Burkhart BR. Race and victim age matter: Sexual behaviors and experiences among confined African American and European American youth with sexual and nonsexual offenses. Sex Abuse 2017 Jul 31 Epub ahead of print (forthcoming). [doi: 10.1177/1079063217720926] [Medline: 28760088]
12. Tomaszewska P, Krahé B. Predictors of sexual aggression victimization and perpetration among Polish university students: a longitudinal study. Arch Sex Behav 2018 Feb;47(2):493-505. [doi: 10.1007/s10508-016-0823-2] [Medline: 27543105]
13. Alquaiz AM, Almuneef MA, Minhas HR. Knowledge, attitudes, and resources of sex education among female adolescents in public and private schools in Central Saudi Arabia. Saudi Med J 2012 Sep;33(9):1001-1009. [Medline: 22964813]
14. Metheny WP, Espey EL, Bienstock J, Cox SM, Erickson SS, Goepfert AR, et al. To the point: medical education reviews evaluation in context: assessing learners, teachers, and training programs. Am J Obstet Gynecol 2005 Jan;192(1):34-37. [doi: 10.1016/j.ajog.2004.07.036] [Medline: 15671999]
15. Milhausen RR, Herold ES. Does the sexual double standard still exist? Perceptions of university women. J Sex Res 1999 Nov;36(4):361-368. [doi: 10.1080/00224499909552008]
16. Holland J, Ramazanoglu C, Sharpe S, Thomson R. The Male in the Head: Young People, Heterosexuality and Power. London: Tufnell Press; 1998.
17. Greene K, Faulkner SL. Gender, belief in the sexual double standard, and sexual talk in heterosexual dating relationships. Sex Roles 2005 Aug;53(3-4):239-251. [doi: 10.1007/s11199-005-5682-6]
18. Marks MJ, Chris Fraley R. The impact of social interaction on the sexual double standard. Soc Influence 2007 Mar;2(1):29-54. [doi: 10.1080/15534510601154413]
19. Tolman DL. Dilemmas of Desire: Teenage Girls Talk about Sexuality. Cambridge, MA: Harvard University Press; 2005.
20. Koenig MA, Zablotska I, Lutalo T, Nalugoda F, Wagman J, Gray R. Coerced first intercourse and reproductive health among adolescent women in Rakai, Uganda. Int Fam Plan Perspect 2004 Dec;30(4):156-163 [FREE Full text] [doi: 10.1363/ifpp.30.156.04] [Medline: 15590381]
21. Schulz JJ, Schulz L. The darkest of ages: Afghan women under the Taliban. Peace Conf 1999;5(3):237-254. [doi: 10.1207/s15327949pac0503_5]
22. Sodhar ZA, Shaikh AG, Sodhar KN. Woman's social rights in Islam: an evaluation of equality of rights between men and women. Grassroots 2015;49(1):171-178.
23. Arfaoui K, Moghadam VM. Violence against women and Tunisian feminism: advocacy, policy, and politics in an Arab context. Curr Sociol 2016 Apr 13;64(4):637-653. [doi: 10.1177/0011392116640481]
24. Moghadam VM. Women and democracy after the Arab Spring: Theory, practice, and prospects. In: Shalaby M, Moghadam VM, editors. Empowering Women after the Arab Spring. New York: Palgrave Macmillan; 2016:193-215.
25. Cooney M. Death by family: Honor violence as punishment. Punishm Soc 2014 Oct;16(4):406-427. [doi: 10.1177/1462474514539537]
26. Flood M. Exposure to pornography among youth in Australia. J Sociol 2007 Mar 01;43(1):45-60. [doi: 10.1177/1440783307073934]
27. Lo V, Wei R. Third-Person Effect, gender, and pornography on the internet. J Broadcast Electron Media 2002 Mar;46(1):13-33. [doi: 10.1207/s15506878jobem4601_2]
28. Cooper A, Boies S, Maheu M, Greenfield D. Sexuality and the internet: the next sexual revolution. In: Psychological perspectives on human sexuality. New York, NY: Wiley; 1999:519-545.
29. Mitchell KJ, Wolak J, Finkelhor D. Trends in youth reports of sexual solicitations, harassment and unwanted exposure to pornography on the internet. J Adolesc Health 2007 Feb;40(2):116-126. [doi: 10.1016/j.jadohealth.2006.05.021] [Medline: 17259051]
30. Wolak J, Mitchell K, Finkelhor D. Unwanted and wanted exposure to online pornography in a national sample of youth internet users. Pediatrics 2007 Feb;119(2):247-257. [doi: 10.1542/peds.2006-1891] [Medline: 17272613]
31. Lenhart A, Ling R, Campbell S, Purcell K. Pew Internet. 2010 Apr 20. Teens and mobile phones: Text messaging explodes as teens embrace it as the centerpiece of their communication strategies with friends URL: http://www.pewinternet.org/2010/04/20/teens-and-mobile-phones/ [accessed 2018-10-01] [WebCite Cache ID 72qVqH0G3]
32. Lenhart A, Purcell K, Smith A, Zickuhr K. Pew Internet. 2010 Feb 03. Social media and young adults URL: http://www.pewinternet.org/2010/02/03/social-media-and-young-adults/ [accessed 2018-10-01] [WebCite Cache ID 72qWGl8t1]

33. Lawsky D. Reuters. 2008 Nov 24. American youth trail in internet use: Survey URL: https://www.reuters.com/article/us-internet-youth/american-youth-trail-in-internet-use-survey-idUSTRE4AN0MR20081124 [accessed 2018-10-01] [WebCite Cache ID 72qWSL1Lx]
34. Livingstone S, Helsper EJ. Taking risks when communicating on the internet: the role of offline social-psychological factors in young people's vulnerability to online risks. Inf Commun Soc 2007 Oct;10(5):619-644. [doi: 10.1080/13691180701657998]
35. Holloway S, Valentine G. Cyberkids: Youth Identities and Communities in an On-Line World. New York, NY: Routledge; 2014.
36. Mesch GS. Social bonds and internet pornographic exposure among adolescents. J Adolesc 2009 Jun;32(3):601-618. [doi: 10.1016/j.adolescence.2008.06.004] [Medline: 18694593]
37. Yen Lai P, Dong Y, Wang M, Wang X. The intervention and regulation of pornography: internal punishment, negative externality, and legal paternalism. J Glob Econ 2014;3(128):2. [doi: 10.4172/2375-4389.1000128]
38. Balmer Jr S. The limits of free speech, pornography and the law. ASLR 2010;1:66.
39. Peter J, Valkenburg PM. Adolescents and pornography: a review of 20 years of research. J Sex Res 2016 Mar;53(4-5):509-531. [doi: 10.1080/00224499.2016.1143441] [Medline: 27105446]
40. Mitchell KJ, Finkelhor D, Wolak J. The exposure of youth To unwanted sexual material on the internet: a national survey of risk, impact, and prevention. Youth Soc 2003 Mar 01;34(3):330-358. [doi: 10.1177/0044118x02250123]
41. Greenfield PM. Inadvertent exposure to pornography on the internet: implications of peer-to-peer file-sharing networks for child development and families. J Appl Dev Psychol 2004 Nov;25(6):741-750. [doi: 10.1016/j.appdev.2004.09.009]
42. Livingstone S, Bober M. London, UK: London School of Economics and Political Science; 2005. UK children go online: Final report of key project findings URL: http://eprints.lse.ac.uk/399/ [accessed 2018-10-01] [WebCite Cache ID 72qXq42ZD]
43. Peter J, Valkenburg PM. Adolescents' exposure to sexually explicit internet material, sexual uncertainty, and attitudes toward uncommitted sexual exploration: is there a link? Commun Res 2008 Aug 04;35(5):579-601. [doi: 10.1177/0093650208321754]
44. Perry LD. 2016 Jun. The impact of pornography on children URL: http://www.acpeds.org/the-college-speaks/position-statements/the-impact-of-pornography-on-children [accessed 2018-10-01] [WebCite Cache ID 72qYCTsHy]
45. Lim MS, Agius PA, Carrotte ER, Vella AM, Hellard ME. Young Australians' use of pornography and associations with sexual risk behaviours. Aust N Z J Public Health 2017 Aug;41(4):438-443. [doi: 10.1111/1753-6405.12678] [Medline: 28664609]
46. Brown BB. "You're going with who?": Peer group influences on adolescent romantic relationships. In: Furman W, Brown BB, Feiring C, editors. The Development of Romantic Relationships in Adolescence. New York, NY: Cambridge University Press; 1999:291-329.
47. Connolly J, Goldberg A. Romantic relationships in adolescence: The role of friends and peers in their emergence and development. In: Furman W, Brown BB, Feiring C, editors. The Development of Romantic Relationships in Adolescence. New York, NY: Cambridge University Press; 1999:266-290.
48. Valkenburg PM, Peter J, Walther JB. Media effects: theory and research. Annu Rev Psychol 2016 Jan;67:315-338. [doi: 10.1146/annurev-psych-122414-033608] [Medline: 26331344]
49. Morrison TG, Ellis SR, Morrison MA, Bearden A, Harriman RL. Exposure to sexually explicit material and variations in body esteem, genital attitudes, and sexual esteem among a sample of Canadian men. J Men's Stud 2007 Mar 1;14(2):209-222. [doi: 10.3149/jms.1402.209]
50. Byrne D, Osland J. Sexual fantasy and erotica/pornography: Internal and external imagery. In: Inzuchman T, Muscarella F, editors. Psychological Perspectives on Human Sexuality. New York, NY: Wiley; 2000:283-305.
51. Mikorski R, Szymanski DM. Masculine norms, peer group, pornography, Facebook, and men’s sexual objectification of women. Psychol Men Masc 2017 Oct;18(4):257-267. [doi: 10.1037/men0000058]
52. Morelli M, Bianchi D, Baiocco R, Pezzuti L, Chirumbolo A. Sexting, psychological distress and dating violence among adolescents and young adults. Psicothema 2016 May;28(2):137-142. [doi: 10.7334/psicothema2015.193] [Medline: 27112809]
53. Bianchi D, Morelli M, Baiocco R, Chirumbolo A. Sexting as the mirror on the wall: body-esteem attribution, media models, and objectified-body consciousness. J Adolesc 2017 Dec;61:164-172. [doi: 10.1016/j.adolescence.2017.10.006] [Medline: 29111447]
54. Seto MC, Maric A, Barbaree HE. The role of pornography in the etiology of sexual aggression. Aggress Violent Behav 2001 Jan;6(1):35-53. [doi: 10.1016/S1359-1789(99)00007-5]
55. Kadri N, Benjelloun R, Kendili I, Khoubila A, Moussaoui D. Internet and sexuality in Morocco, from cyber habits to psychopathology. Sexologies 2013 Apr;22(2):e49-e53. [doi: 10.1016/j.sexol.2012.08.006]
56. Brombers M, Theokas C. Washington, DC; 2013 May. Breaking the glass ceiling of achievement for low-income students and students of color URL: https://files.eric.ed.gov/fulltext/ED543218.pdf [accessed 2018-10-01] [WebCite Cache ID 72qkPmvcm]
57. Kasemy Z, Desouky DE, Abdelrasoul G. Sexual fantasy, masturbation and pornography among Egyptians. Sex Cult 2016 Mar 12;20(3):626-638. [doi: 10.1007/s12119-016-9346-1]
58. Hawi NS. Internet addiction among adolescents in Lebanon. Comput Human Behav 2012 May;28(3):1044-1053. [doi: 10.1016/j.chb.2012.01.007]

59. Owens EW, Behun RJ, Manning JC, Reid RC. The impact of internet pornography on adolescents: a review of the research. Sex Addict Compulsivity 2012 Jan;19(1-2):99-122. [doi: 10.1080/10720162.2012.660431]
60. Creswell JW, Hanson WE, Clark Plano VL, Morales A. Qualitative research designs: selection and implementation. Couns Psychol 2007 Mar 01;35(2):236-264. [doi: 10.1177/0011000006287390]
61. Parent AS, Teilmann G, Juul A, Skakkebaek NE, Toppari J, Bourguignon JP. The timing of normal puberty and the age limits of sexual precocity: variations around the world, secular trends, and changes after migration. Endocr Rev 2003 Oct;24(5):668-693. [doi: 10.1210/er.2002-0019] [Medline: 14570750]
62. 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]
63. Corbin J, Strauss A. Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory. 4 Edition. Thousand Oaks, CA: Sage; 2015.
64. Roudi-Fahimi F. Washington, DC; 2003 Feb. Women's reproductive health in the Middle East and North Africa URL: https://www.prb.org/womensreproductivehealthinthemiddleeastandnorthafricapdf234kb/ [accessed 2018-10-01] [WebCite Cache ID 72qlTSwuc]
65. Erez E, Ibarra PR, Gur OM. At the intersection of private and political conflict zones: policing domestic violence in the Arab community in Israel. Int J Offender Ther Comp Criminol 2015 Aug;59(9):930-963. [doi: 10.1177/0306624X14532602] [Medline: 24831799]
66. Yasmine R, El Salibi N, El Kak F, Ghandour L. Postponing sexual debut among university youth: how do men and women differ in their perceptions, values and non-penetrative sexual practices? Cult Health Sex 2015;17(5):555-575. [doi: 10.1080/13691058.2014.972457]
67. Allen L. Girls want sex, boys want love: resisting dominant discourses of (hetero) sexuality. Sexualities 2003 May 11;6(2):215-236. [doi: 10.1177/1363460703006002004]
68. Baron-Epel O, Bord S, Elias W, Zarecki C, Shiftan Y, Gesser-Edelsburg A. Alcohol consumption among Arabs in Israel: a qualitative study. Subst Use Misuse 2015 Jan;50(2):268-273. [doi: 10.3109/10826084.2014.962051] [Medline: 25295595]
69. Falligant JM, Alexander AA, Burkhart BR. Risk assessment of juveniles adjudicated for possession of child sexual exploitation material. J Forensic Psychol Pract 2017 Feb 16;17(2):145-156. [doi: 10.1080/15228932.2017.1270640]

Abbreviations

STD: sexually transmitted disease

Edited by G Eysenbach; submitted 23.07.18; peer-reviewed by G Mesch, N Amoyal Pensak; comments to author 16.08.18; revised version received 10.09.18; accepted 22.09.18; published 09.10.18.

Please cite as:
Gesser-Edelsburg A, Abed Elhadi Arabia M
Discourse on Exposure to Pornography Content Online Between Arab Adolescents and Parents: Qualitative Study on its Impact on Sexual Education and Behavior
J Med Internet Res 2018;20(10):e11667
URL: http://www.jmir.org/2018/10/e11667/
doi:10.2196/11667
PMID:30305264

Original Paper

Assessment of Personal Health Care Management and Chronic Disease Prevalence: Comparative Analysis of Demographic, Socioeconomic, and Health-Related Variables

Ryan H Sandefer[1], PhD; Bonnie L Westra[2], RN, PhD, FAAN, FACMI; Saif S Khairat[3,4], PhD; David S Pieczkiewicz[5], PhD; Stuart M Speedie[5], PhD, FACMI

[1]Department of Health Informatics and Information Management, College of St. Scholastica, Duluth, MN, United States
[2]Center for Nursing Informatics, School of Nursing & Institute for Health Informatics, University of Minnesota, Minneapolis, MN, United States
[3]Carolina Health Informatics Program, 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
[5]Institute for Health Informatics, University of Minnesota Twin Cities, Minneapolis, MN, United States

Corresponding Author:
Ryan H Sandefer, PhD
Department of Health Informatics and Information Management
College of St. Scholastica
1200 Kenwood Avenue
Duluth, MN, 55811
United States
Phone: 1 218 625 4931
Email: rsandefe@css.edu

Abstract

Background: The use of personal health care management (PHM) is increasing rapidly within the United States because of implementation of health technology across the health care continuum and increased regulatory requirements for health care providers and organizations promoting the use of PHM, particularly the use of text messaging (short message service), Web-based scheduling, and Web-based requests for prescription renewals. Limited research has been conducted comparing PHM use across groups based on chronic conditions.

Objective: This study aimed to describe the overall utilization of PHM and compare individual characteristics associated with PHM in groups with no reported chronic conditions, with 1 chronic condition, and with 2 or more such conditions.

Methods: Datasets drawn from the National Health Interview Survey were analyzed using multiple logistic regression to determine the level of PHM use in relation to demographic, socioeconomic, or health-related factors. Data from 47,814 individuals were analyzed using logistic regression.

Results: Approximately 12.19% (5737/47,814) of respondents reported using PHM, but higher rates of use were reported by individuals with higher levels of education and income. The overall rate of PHM remained stable between 2009 and 2014, despite increased focus on the promotion of patient engagement initiatives. Demographic factors predictive of PHM use included people who were younger, non-Hispanic, and who lived in the western region of the United States. There were also differences in PHM use based on socioeconomic factors. Respondents with college-level education were over 2.5 times more likely to use PHM than respondents without college-level education. Health-related factors were also predictive of PHM use. Individuals with health insurance and a usual place for health care were more likely to use PHM than individuals with no health insurance and no usual place for health care. Individuals reporting a single chronic condition or multiple chronic conditions reported slightly higher levels of PHM use than individuals reporting no chronic conditions. Individuals with no chronic conditions who did not experience barriers to accessing health care were more likely to use PHM than individuals with 1 or more chronic conditions.

Conclusions: The findings of this study illustrated the disparities in PHM use based on the number of chronic conditions and that multiple factors influence the use of PHM, including economics and education. These findings provide evidence of the challenge associated with engaging patients using electronic health information as the health care industry continues to evolve.

(J Med Internet Res 2018;20(10):e276) doi:10.2196/jmir.8784

KEYWORDS

personal health information; consumer participation; social determinants of health; personal health records

Introduction

Background

Patient access to their electronic health information has been identified as a key priority for improving care quality and efficiency [1,2]. Individual access to and personal use of health information is a cornerstone of recent national health care efforts. As quoted in a study, “With access to their electronic health information, individuals can serve as intermediaries of information exchange among providers and use innovative applications to better manage their health” [3]. The Centers for Medicare and Medicaid Services (CMS) has incorporated electronic access to health information within the electronic health record (EHR) Incentive Program, which requires eligible professionals and hospitals to demonstrate meaningful use of EHR systems. Patient and family engagement is 1 of the 4 primary goals of the EHR Incentive Program, and the program includes multiple measures of patient and family engagement through the use of health information technology. These include sending and receiving secure messages between patients and providers; providing Web-based access to view, download, and transmit health information; and identifying and providing patient-specific educational resources based on clinically relevant information housed within the certified EHR system [1]. The CMS accountable care organization also promotes patient engagement and care coordination in an effort to restrain costs, improve patient experience of care, improve self-management, and facilitate communication between patients and providers [4].

These federal programs require measures of patient and family engagement because of the impact of the individual use of health information on care quality. Multiple studies have identified an association between personal use of health information and improvements in chronic disease management [5-9]. Studies have also shown an association between personal use of health information and improvements in clinical quality outcomes, patient satisfaction, and overall perceptions regarding the efficacy of communication between patients and health care stakeholders [10-15].

In an effort to meet the program goals of the EHR Incentive Program and accountable care organizations, US health care organizations have started implementing EHRs, patient portals, and personal health records at an unprecedented rate. The number of US nonfederal acute care hospitals with the capability of offering patients the ability to electronically view, download, and transmit their health information increased from 10% in 2012 to 69% in 2015. Over 95% of US hospitals provide patients the ability to view their health information electronically [16]. Similarly, between 2013 and 2014, the number of individuals who were offered access to their Web-based medical records increased from 28% to 38%. Over half of the individuals who were offered access to their Web-based medical record in 2014 accessed it [17].

The focus on electronic access to patient health information is not limited to the United States. There are international efforts to provide access to Web-based health information to address issues of access, affordability, and quality. Research on adoption and use of these systems has been conducted in Denmark, Canada, Australia, and Estonia, among others [18-21].

Despite the rapid increase in the capability to view, download, and transmit personal health information, there are disparities regarding individual access and use of their health information. According to one estimate, approximately 4 of 10 US adults used some type of health information technology in 2013, but individuals with less education, lower incomes, or those who lived in rural areas were less likely to email health care providers (HCPs), view laboratory results on the Web, and access health-related information with mobile phones [3]. Similar research has demonstrated disparities in access and use of electronic health information [22-24]. Use of technology for managing personal health information is associated with age [23-27], race [7,24,25,28,29], ethnicity [23-25], and gender [7,23,25]. The socioeconomic factors of income and education level are also related to an individual’s use of technology in accessing and using health information [3,7,24,25,28].

Individuals who report having chronic conditions are more likely to electronically access and use personal health information and are also more likely to access personal health information repeatedly [24,28,30,31]. Nearly half of the US population suffers from at least 1 chronic condition, and nearly 12% of the US population reports having 3 or more chronic conditions [32]. The advantage of using technologies to electronically access health information is more effective management and coordination of care [33]. Individuals managing 1 or more chronic conditions, likely across numerous HCPs and institutions, have the potential to benefit by accessing their information electronically, downloading the information, and sharing it with other members of the care team.

Objectives

Although individuals with chronic health conditions appear to access electronic health information more frequently than others, there has been little research conducted regarding the relationship between the number of chronic conditions and participation in electronic personal health care management (PHM). In this project, we are interested in how demographic, socioeconomic, and health-related variables affect PHM. PHM, as defined here, refers to the individual use of internet-based technology to access personal health information or communicate with HCPs regarding patient health information. PHM is considered active participation with a health care entity through the use of technology, and PHM refers to use of technology-mediated apps by an individual to assist in meeting her or his health care–related needs. The purpose of this research was to describe the overall utilization of PHM and compare individual characteristics associated with PHM among groups with no reported chronic conditions, with 1 chronic condition, and with 2 or more such conditions.

Methods

Sample

Data from the US National Health Interview Survey (NHIS) aggregated by the Integrated Health Interview Series (IHIS) were used for this analysis. The IHIS has collected and harmonized 52 years of NHIS data for the purpose of research and analysis [33]. The Centers for Disease Control and Prevention annually conducts the NHIS “to secure accurate and current statistical information on the amount, distribution, and effects of illness and disability in the United States and the services rendered for or because of such conditions.” The NHIS sample is representative of the US population drawn from each US state and the District of Columbia and includes approximately 35,000 households and 87,500 persons annually. It has an average response rate of approximately 90%, and it has been conducted annually since 1957 [34]. Multiple publications have detailed the use of the NHIS [35,36]. For the purpose of this project, an SAS text file was downloaded from the IHIS portal and imported into the R statistical software package for analysis (R Foundation for Statistical Computing).

NHIS survey data regarding adults older than 18 years from 2009, 2011, 2012, 2013, to 2014 were combined for this study. The 2010 survey was excluded because it did not collect the information regarding PHM required for this analysis. Between 2009 and 2014, there were 605,001 individuals interviewed. The data used in this study were limited to adults who were asked and responded “Yes” or “No” to specific questions regarding PHM and were further limited to only individuals with complete data for the variables included in the analysis, resulting in a study sample size of 50,062 individuals.

Variable Selection

The NHIS includes items related to a variety of health care–related characteristics, demographics, health conditions, and behaviors. A dichotomous variable labeled PHM was calculated based on the questionnaire items related to text messaging (short message service) HCPs, refilling of prescriptions on the Web, and scheduling of health care appointments on the Web. PHM indicates use of electronic health information management. The levels of responses for each question were “Yes,” “No,” “Refused,” “Not Ascertained,” or “Don’t Know.” A response of “Yes” to any of the 3 questions resulted in a “Yes” PHM response, otherwise a “No” was assigned. For the purpose of the analysis, “No” was coded as 0 and “Yes” was coded as 1. Any response of “Refused,” “Don’t Know,” or “Not Ascertained” to any of the 3 questions resulted in the removal of that respondent from the analysis. Individual use of PHM was the dependent variable. The entire list of predictor variables and their assignable value sets are listed in the Results section.

Statistical Analysis

The R statistical software package, version 3.2.3, was used for all statistical analyses. The *survey* package allows the analysis to account for complexity of the NHIS sample. To account for combining multiple years of NHIS data, the sample weight was divided by the number of years (5 years) of data included in the analysis [37]. Multiple logistic regression was used to identify predictors of PHM among multiple independent samples based on chronic condition status.

Moreover, 3 different models of PHM were created to characterize and compare the 3 subgroups of respondents. The first model included only respondents who did not report any of the 5 chronic conditions (diabetes, hypertension, asthma, heart condition, or arthritis) and represented the ability to predict PHM for individuals without chronic conditions. The second model included only individuals reported as having 1 of the 5 chronic conditions (one chronic condition). The third model included all observations where respondents reported having 2 or more of the 5 chronic conditions included in this study (multiple chronic conditions).

Descriptive statistics for each group and each predictor variable were also calculated. A 2-sample test for equality of proportions was used to compare PHM use.

Protection of Human and Animal Subjects

Analysis of the NHIS data was deemed exempt from review by the University of Minnesota Institutional Review Board.

Results

Respondent Characteristics

The total NHIS sample for this study after exclusions was 50,061 individuals—the number of total responses without controlling for missing values was 75,305 with 8082 positive PHM responses (10.73%, 8082/75,305). The NHIS sample was separated into 3 mutually exclusive subgroups based on the number of reported chronic conditions. The sizes of the samples were as follows: 22,965 (no chronic condition), 13,325 (1 chronic condition), and 13,771 (multiple chronic conditions).

Overall Personal Health Care Management Use

The overall proportion of US adults who reported PHM use between 2009 and 2014 was 12.19%, and the use of PHM has increased slightly over that period; Figure 1 provides a breakdown by group and by year. Overall, there was an increase in reported PHM use from 11.24% to 13.27% between 2012 and 2013.

PHM use for each subgroup characterized by demographic, socioeconomic, and health-related characteristics are listed in Tables 1-3.

The overall level of use of PHM varied slightly by group. The highest overall rates of PHM use were among individuals reporting 1 chronic condition. The results show that 14.97% (1996/13,325) individuals with a single chronic condition reported PHM use, followed by 14.73% (2029/13,771) with multiple chronic conditions and 11.5% (2632/22,965) with no chronic conditions. The proportion of PHM use was significantly higher for individuals reporting either a single chronic condition or multiple chronic conditions than those with no chronic conditions (*P*<.001). There was no difference between the proportion of PHM use between those who reported a single chronic condition or multiple chronic conditions (*P*=.84).

Figure 1. Proportion of US adults using personal health management (PHM) by year by chronic condition group. CC: chronic condition.

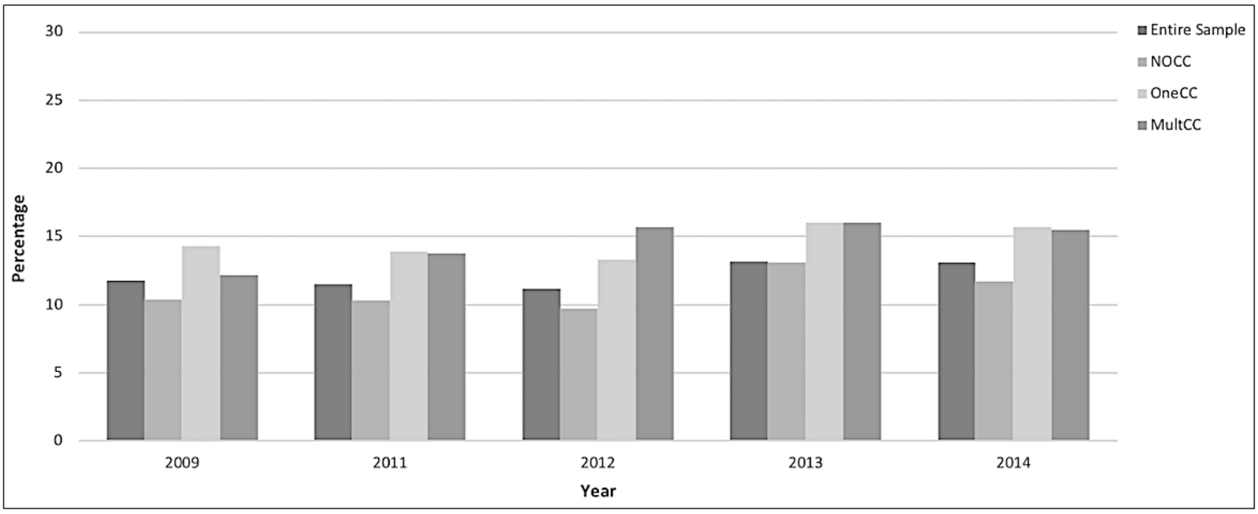


Table 1. Demographic characteristics of respondents who reported personal health management (PHM) use by number of chronic conditions.

Variable	Without chronic condition (N=22,965)		One chronic condition (N=13,325)		Multiple chronic conditions (N=13,771)	
	All, n (%)	PHM, n (%)	All, n (%)	PHM, n (%)	All, n (%)	PHM, n (%)
Age in years						
18-40	13,687 (59.60)	1644 (12.01)	4476 (33.59)	801 (17.90)	1337 (9.71)	260 (19.45)
41-60	7233 (31.50)	816 (11.28)	5213 (39.12)	797 (15.29)	4665 (33.88)	846 (18.14)
60+	2045 (8.90)	172 (8.41)	3636 (27.29)	398 (10.95)	7769 (56.42)	923 (11.88)
Sex						
Male	10,830 (47.16)	919 (8.49)	6065 (45.52)	775 (12.78)	5733 (41.63)	902 (15.73)
Female	12,135 (52.84)	1713 (14.12)	7260 (54.48)	1221 (16.82)	8038 (58.37)	1127 (14.02)
Race						
White	17,427 (75.89)	2049 (11.76)	10,271 (77.08)	1593 (15.51)	10,541 (76.54)	1686 (15.99)
Nonwhite	5538 (24.11)	583 (10.53)	3054 (22.92)	403 (13.20)	3230 (23.46)	343 (10.62)
Ethnicity						
Not Hispanic	17,863 (77.78)	2268 (12.70)	11,442 (85.87)	1827 (15.97)	12,234 (88.84)	1904 (15.56)
Hispanic	5102 (22.22)	364 (7.13)	1883 (14.13)	169 (8.98)	1537 (11.16)	125 (8.13)
Born in United States						
Yes	17,272 (75.21)	2147 (12.43)	11,297 (84.78)	1769 (15.66)	12,155 (88.27)	1907 (15.69)
No	5693 (24.79)	485 (8.52)	2028 (15.22)	227 (11.19)	1616 (11.73)	122 (7.55)
Geography						
Midwest	4864 (21.18)	538 (11.06)	2894 (21.72)	408 (14.10)	2961 (21.50)	411 (13.88)
Northeast	3553 (15.47)	369 (10.39)	2165 (16.25)	284 (13.12)	2249 (16.33)	316 (14.05)
South	8032 (34.97)	829 (10.32)	4800 (36.02)	684 (14.25)	5314 (38.59)	679 (12.78)
West	6516 (28.37)	896 (13.75)	3466 (26.01)	620 (17.89)	3247 (23.58)	623 (19.19)

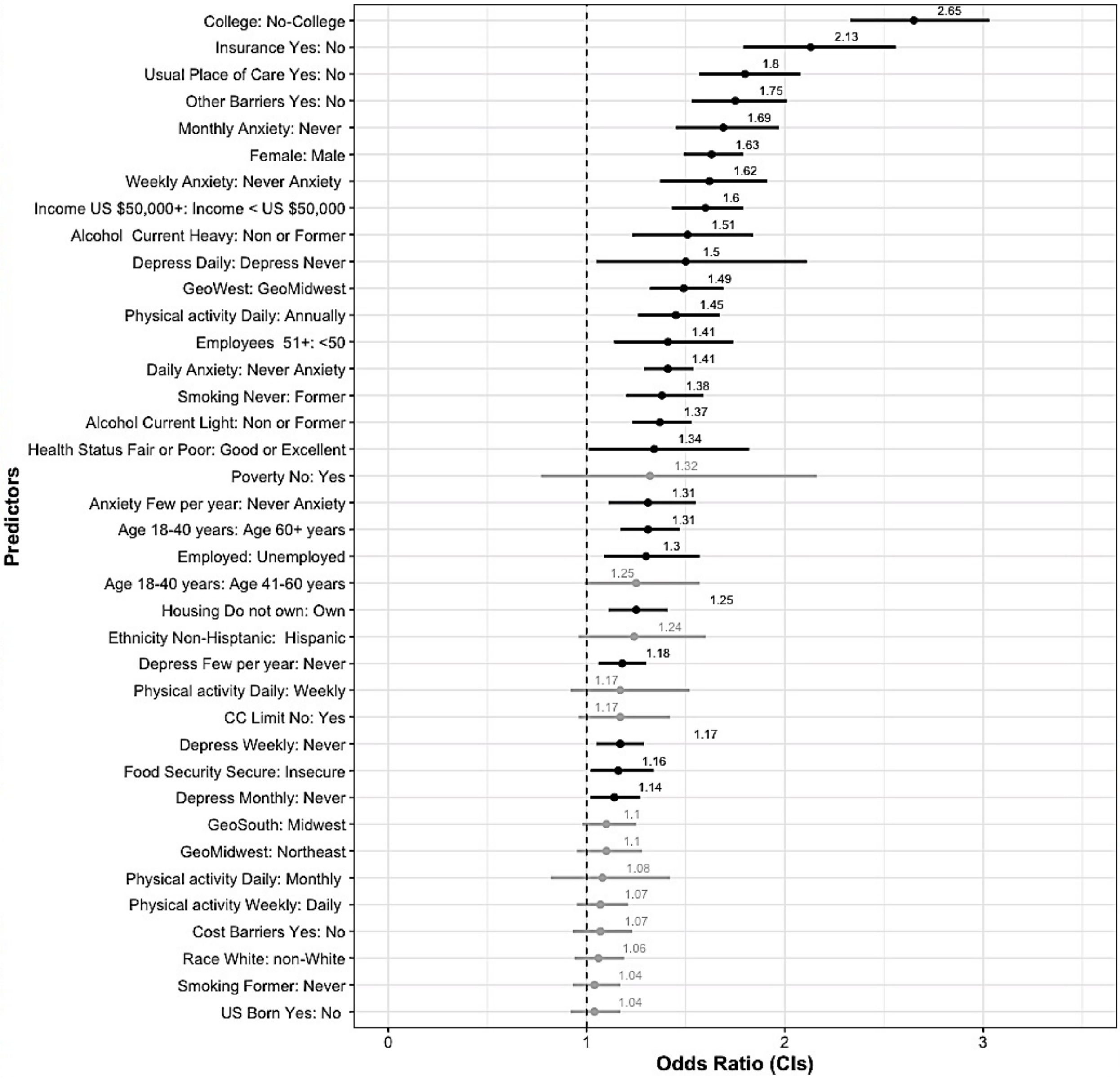
Table 2. Socioeconomic characteristics of respondents who reported personal health management (PHM) use by number of chronic conditions.

Variable	Without chronic condition (N=22,965)		One chronic condition (N=13,325)		Multiple chronic conditions (N=13,771)	
	All, n (%)	PHM, n (%)	All, n (%)	PHM, n (%)	All, n (%)	PHM, (%)
Education						
No college	8443 (36.76)	318 (3.77)	5133 (38.52)	268 (5.22)	6495 (47.16)	386 (5.94)
College	14,522 (63.24)	2314 (15.93)	8192 (61.48)	1728 (21.09)	7276 (52.84)	1643 (22.58)
Family income (US $)						
<50,000	12,023 (52.35)	844 (7.02)	7053 (52.93)	639 (9.06)	8633 (62.69)	733 (8.49)
50,000+	10,942 (47.65)	1788 (16.34)	6272 (47.07)	1357 (21.64)	5138 (37.31)	1296 (25.22)
Poverty						
Yes	4210 (18.33)	224 (5.32)	2218 (16.65)	148 (6.67)	2500 (18.15)	131 (5.24)
No	18,755 (81.67)	2408 (12.84)	11,107 (83.35)	1848 (16.64)	11,271 (81.85)	1898 (16.84)
Number of employees						
<50	15,302 (66.63)	1421 (9.29)	8467 (63.54)	1076 (12.71)	8442 (61.30)	1081 (12.81)
51+	7663 (33.37)	1211 (15.80)	4858 (36.46)	920 (18.94)	5329 (38.70)	948 (17.79)
Employed						
Employed	16,634 (72.43)	2139 (12.86)	8063 (60.51)	1448 (17.96)	4984 (36.19)	1093 (21.93)
Unemployed	6331 (27.57)	493 (7.79)	5262 (39.49)	548 (10.41)	8787 (63.80)	936 (10.65)
Insurance						
Yes	17,888 (77.89)	2462 (13.76)	11,308 (84.86)	1868 (16.52)	12,716 (92.34)	1950 (15.34)
No	5077 (22.11)	170 (3.35)	2017 (15.14)	128 (6.35)	1055 (7.66)	79 (7.49)
Housing						
Own home	12,024 (52.36)	1533 (12.75)	8231 (61.77)	1339 (16.27)	9057 (65.88)	1493 (16.48)
Do not own	10,941 (47.64)	1099 (10.04)	5094 (38.23)	657 (12.90)	4714 (34.23)	536 (11.37)
Food security						
Insecure	1291 (5.62)	84 (6.51)	991 (7.44)	69 (6.96)	1459 (10.59)	138 (9.46)
Secure	21,674 (94.38)	2548 (11.76)	12,334 (92.56)	1927 (15.62)	12,312 (89.41)	1891 (15.36)
Cost barriers						
Yes	3811 (16.59)	355 (9.32)	2672 (20.05)	342 (12.80)	2678 (19.45)	362 (13.52)
No	19,154 (83.41)	2277 (11.89)	10,653 (79.95)	1654 (15.53)	11,093 (80.55)	1667 (15.03)
Other barriers						
Yes	1825 (7.95)	356 (19.51)	1528 (11.47)	292 (19.11)	11,618 (84.37)	1666 (14.34)
No	21,140 (92.05)	2276 (10.77)	11,797 (88.53)	1704 (14.44)	2153 (15.63)	363 (16.86)

Table 3. Health-related characteristics of respondents who reported personal health management (PHM) use by number of chronic conditions.

Variable	Without chronic condition (N=22,965)		One chronic condition (N=13,325)		Multiple chronic conditions (N=13,771)	
	All, n (%)	PHM, n (%)	All, n (%)	PHM, n (%)	All, n (%)	PHM, n (%)
Health status						
Fair or poor	1050 (4.57)	59 (5.62)	1727 (12.96)	159 (9.21)	4472 (32.47)	458 (10.24)
Good or excellent	21,915 (95.43)	2573 (11.74)	11,598 (87.04)	1837 (15.84)	9299 (67.53)	1571 (16.89)
Usual place of care						
Yes	17,647 (76.84)	2365 (13.40)	11,736 (88.08)	1868 (15.92)	13,179 (95.70)	1980 (15.02)
No	5318 (23.16)	267 (5.02)	1589 (11.92)	128 (8.06)	592 (4.30)	49 (8.28)
Alcohol						
Non or former	7142 (31.10.)	527 (7.38)	4372 (32.81)	388 (8.87)	6192 (44.96)	579 (9.35)
Current light	14,500 (63.14)	1928 (13.30)	8143 (61.11)	1480 (18.18)	6888 (50.02)	1326 (19.25)
Current heavy	1323 (5.76)	177 (13.38)	810 (6.08)	128 (15.80)	691 (5.02)	124 (17.95)
Smoking						
Never	15,308 (66.66)	1876 (12.26)	7500 (56.29)	1262 (16.83)	6791 (49.31)	1045 (15.39)
Former	3478 (15.14)	465 (13.37)	3042 (22.83)	501 (16.47)	4559 (33.11)	741 (16.25)
Current	4179 (18.20)	291 (6.96)	2783 (20.89)	233 (8.37)	2421 (17.58)	243 (10.04)
Limits from chronic conditions						
Yes	1168 (5.09)	115 (9.85)	2160 (16.21)	231 (10.69)	5537 (40.21)	636 (11.49)
No	21,797 (94.91)	2517 (11.55)	11,165 (83.79)	1765 (15.81)	8234 (59.79)	1393 (16.92)
Depression						
Never	15,110 (65.80)	1510 (9.99)	7725 (57.97)	1077 (13.94)	7075 (51.38)	991 (14.01)
Few times per year	5312 (23.13)	778 (14.65)	3497 (26.24)	608 (17.39)	3481 (25.28)	592 (17.01)
Monthly	1202 (5.23)	185 (15.39)	843 (6.33)	152 (18.03)	927 (6.73)	153 (16.50)
Weekly	826 (3.60)	105 (12.71)	688 (5.16)	90 (13.08)	1108 (8.05)	152 (13.72)
Daily	515 (2.24)	54 (10.49)	572 (4.29)	69 (12.06)	1180 (8.57)	141 (11.95)
Anxiety						
Never	10,306 (44.88)	809 (7.85)	5305 (39.81)	621 (11.71)	5007 (36.36)	607 (12.12)
Few times per year	6919 (30.13)	897 (12.96)	3999 (30.01)	654 (16.35)	4038 (29.32)	668 (16.54)
Monthly	2175 (9.47)	387 (17.79)	1297 (9.73)	264 (20.35)	1173 (8.52)	189 (16.11)
Weekly	2104 (9.16)	356 (16.92)	1465 (10.99)	282 (19.25)	1621 (11.77)	284 (17.52)
Daily	1461 (6.36)	183 (12.53)	1259 (9.45)	175 (13.90)	1932 (14.03)	281 (14.54)
Moderate physical activity level						
Daily	3433 (14.95)	423 (12.32)	1978 (14.84)	322 (16.28)	1993 (14.47)	324 (16.26)
Weekly	10,258 (44.67)	1557 (15.18)	5737 (43.05)	1128 (19.66)	4876 (35.41)	1006 (20.63)
Monthly	605 (2.63)	78 (12.89)	361 (2.71)	59 (16.34)	365 (2.65)	63 (17.26)
Annually	131 (0.57)	20 (15.27)	77 (0.58)	12 (15.58)	83 (0.60)	10 (12.05)
Never	8538 (37.18)	554 (6.49)	5172 (38.81)	475 (9.18)	6454 (46.87)	626 (9.70)

Figure 2. No chronic condition (CC) forest plot—multivariate logistic regression odds ratios (ORs) and 95% CIs for all predictor variables ordered by magnitude. Significant ORs are bolded. Reference categories for each predictor are on the right side of the colon.



Personal Health Care Management Use Among Different Groups and Factors That Predict It

Figures 2-4 show the results of the logistic regression analyses, where PHM use was the dependent variable and demographic, socioeconomic, and health-related indicators were the predictors for the groups of respondents.

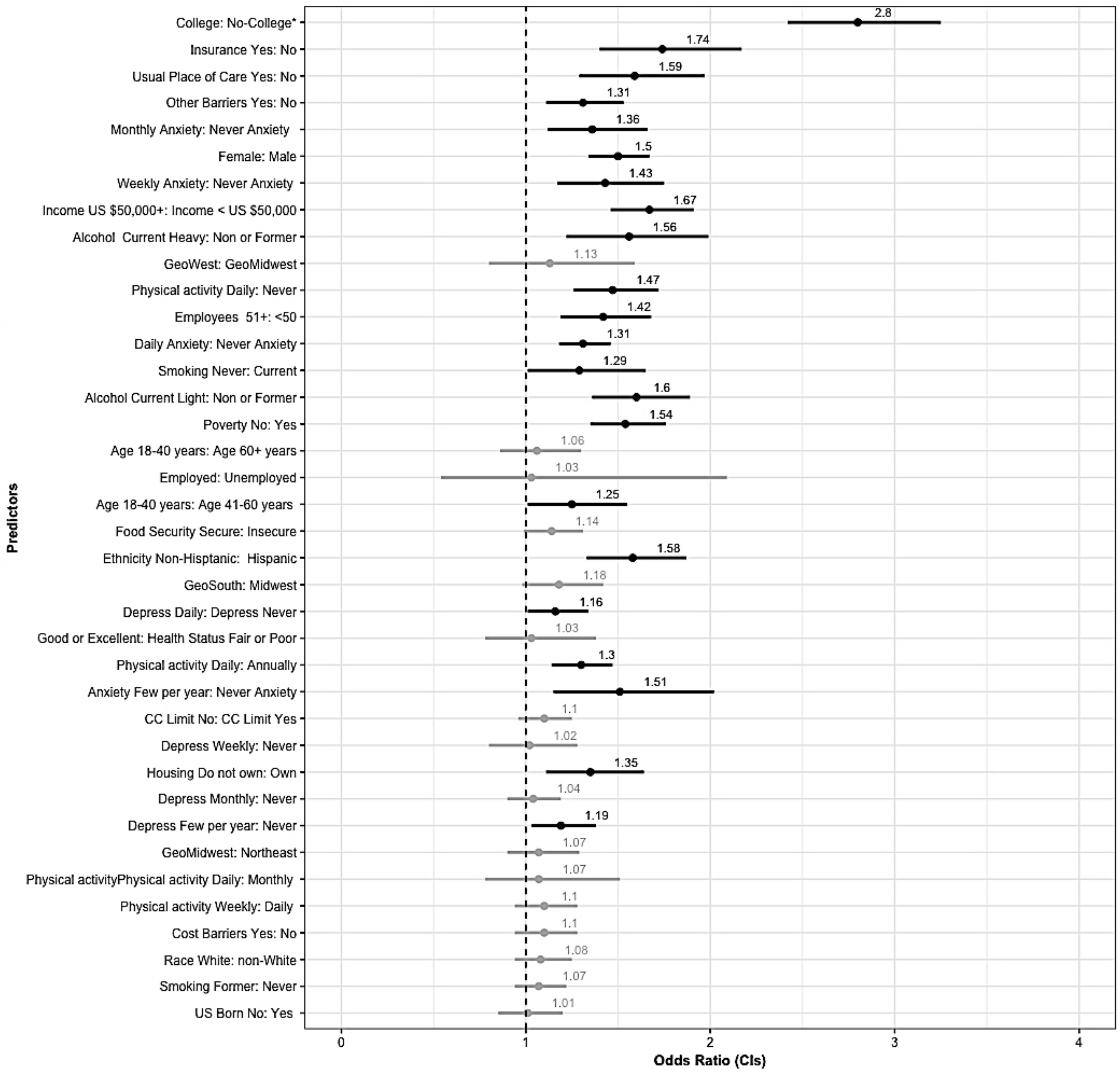
No Reported Chronic Conditions Subgroup

The lowest level of PHM use was from individuals reporting no chronic conditions. Among all factors reported in Tables 1-3, the highest proportional use of PHM was among individuals who reported monthly (17.79%, 387/2175 or weekly (16.92%, 356/2104) anxiety. Those without health insurance had the lowest proportional use of PHM (3.34%, 170/5077), followed by those without college-level education (3.76%, 318/8443) and those in poverty (5.79%, 224/4210). Educational attainment was the factor with the greatest difference in PHM use between levels. Of individuals reporting college-level education, 15.93% (2314/14522) used PHM, whereas only 3.76% (318/8443) of individuals without college-level education reported PHM use.

Figure 2 displays the odds ratios (ORs) and CIs for the variables used in predicting PHM use for respondents reporting no chronic conditions. Interpretations of the ORs using the term “likelihood” or “likely” explicitly refers to a comparison of odds used to calculate the OR for each significant variable. For this group, individuals with college-level education were over 2.6 times more likely to use PHM than those without a college-level education (OR 2.58, 95% CI 2.23-3.0). Respondents with health insurance were over 2 times more likely to use PHM than those without insurance (OR 2.11, 95% CI 1.74-2.54). Those reporting a usual place of care were nearly twice as likely to use PHM than those without a usual place of care (OR 1.81, 95% CI 1.55-2.1), and individuals reporting no other barriers to accessing health care were more likely to use

PHM than those reporting such barriers (OR 1.73, 95% CI 1.49-2.02).

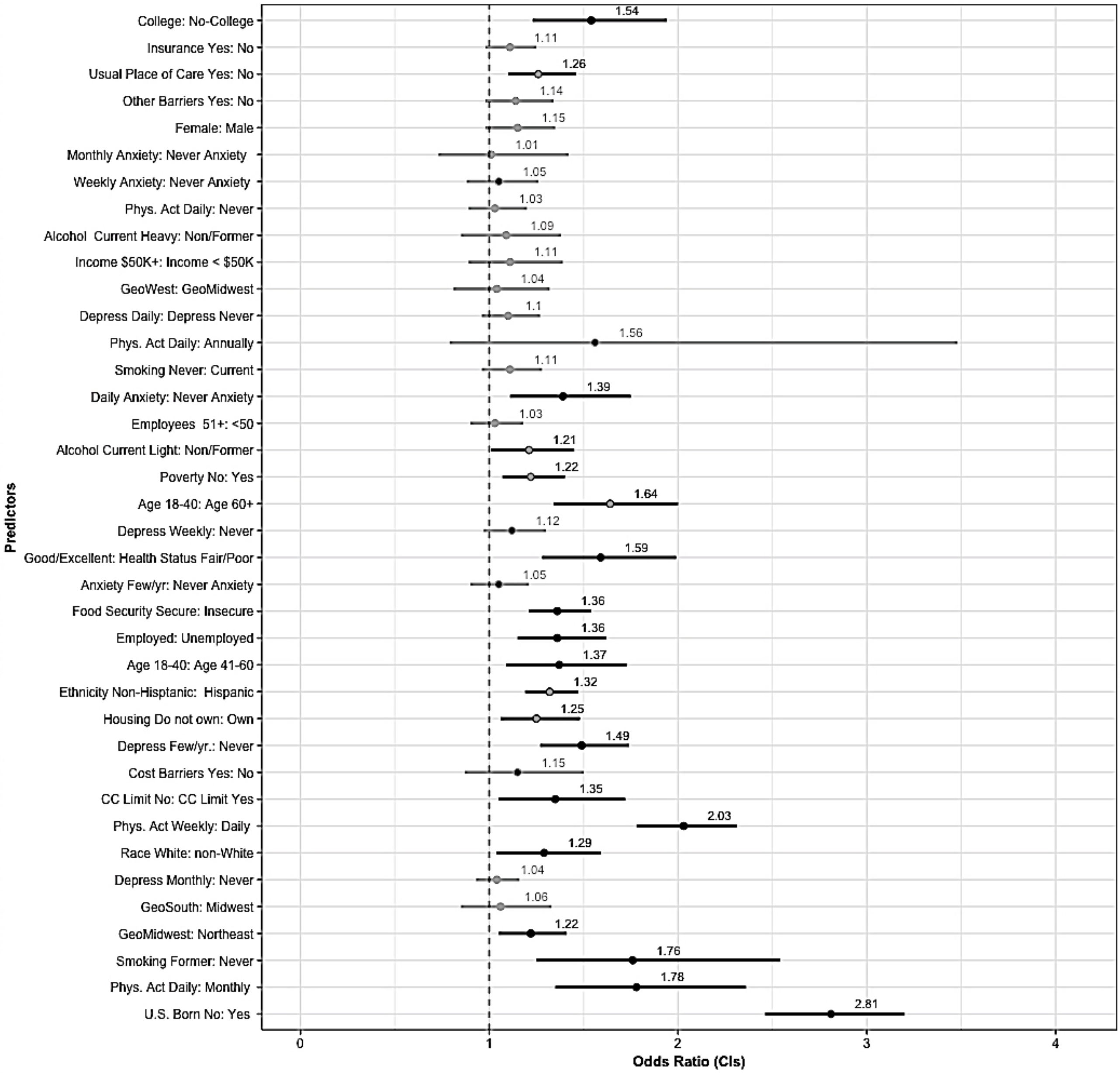
Figure 3. Single chronic condition (CC) forest plot—multivariate logistic regression odds ratios (ORs) and 95% CIs for all predictor variables ordered by magnitude of no chronic condition group. Significant ORs are bolded. Reference categories for each predictor are on the right side of the colon.



Women (OR 1.67, 95% CI 1.51-1.83) and individuals reporting higher family incomes (OR 1.56, 95% CI 1.37-1.77) were also more likely to use PHM. PHM use was higher for those individuals who reported having anxiety a few times per year (OR 1.28, 95% CI 1.13-1.45), having anxiety monthly (OR 1.65, 95% CI 1.39-1.96), having anxiety weekly (OR 1.58, 95% CI 1.29-1.94), and having anxiety daily (OR 1.43, 95% CI 1.12-1.84) and were more likely to use PHM than those who reported no anxiety. Individuals living in the west were more likely to use PHM than individuals in the Midwest (OR 1.52, 95% CI 1.31-1.76).

Individuals who reported engaging in moderate physical activity (PA) daily were nearly one and a half times as likely to use PHM as individuals who reported never engaging in moderate PA (OR 1.48, 95% CI 1.29-1.7). Individuals who reported working at organizations with more than 50 employees were more likely to use PHM than those working in organizations with fewer employees (OR 1.43, 95% CI 1.3-1.57), and those who reported being employed were more likely to use PHM than those who reported being unemployed (OR 1.20, 95% CI 1.03-1.38). Respondents who indicated current heavy (OR 1.56, 95% CI 1.24-1.96) or current light (OR 1.38, 95% CI 1.2-1.58) alcohol consumption were more likely to use PHM than individuals who reported not currently consuming alcohol, and individuals who reported not ever smoking were more likely to use PHM than those who reported currently smoking (OR 1.46, 95% CI 1.23-1.72).

Figure 4. Multiple chronic conditions (CCs) forest plot—multivariate logistic regression odds ratios (ORs) and 95% CIs for all predictor variables ordered by magnitude of the no chronic condition group. Significant ORs are bolded. Reference categories for each predictor are on the right side of the colon.



Single Chronic Condition Subgroup

PHM use for individuals reporting a single chronic condition was 14.97% (1996/13325). Among all factors included in the analysis (see Tables 1-3), the highest proportional use of PHM in this group was among individuals who reported incomes greater than US $50,000 per year (21.63%, 1357/6272), college-level education (21.09%, 1728/8192), and moderate PA weekly (19.66%, 1128/5737). Those without college-level education (5.22%, 268/5133), without health insurance (16.51%, 1868/11,308), and reporting being food insecure (6.96%, 69/991) had the lowest proportional use of PHM. Education level was the factor with the greatest difference in PHM use between levels. Overall, 21% of individuals reporting college-level education used PHM, whereas only 5.2% of individuals without college-level education reported PHM use.

Figure 3 illustrates the predictors of PHM use among individuals reporting a single chronic condition and is ordered by the magnitude of the no chronic condition group. Respondents with a single chronic condition who reported having a college-level education had an increased likelihood of using PHM compared with those without college-level education (OR 2.88, 95% CI 2.44-3.4). Individuals who reported having insurance (OR 1.61, 95% CI 1.26-2.06), a median family income greater than US $50 per year (OR 1.72, 95% CI 1.48-1.99), never smoking (OR 1.75, 95% CI 1.45-2.12), and having a usual place of care (OR 1.72, 95% CI 1.37-2.15) were more likely to use PHM than those who reported no health insurance, median family income of less than US $50 per year, currently smoking, and not having a usual place of health care. Adults aged 18 to 40 years were more likely to use PHM compared with adults aged 60 years and older (OR 1.59, 95% CI 1.32-1.91) and those aged 41 to

60 years (OR 1.35, 95% CI 1.18-1.54). Respondents who reported an alcohol status of current heavy (OR 1.56, 95% CI 1.32-1.84) or current light (OR 1.77, 95% CI 1.36-2.3) were more likely to use PHM than those who reported no alcohol consumption. Individuals who reported being food secure were more likely to use PHM than those who reported being food insecure (OR 1.6, 95% CI 1.16-2.0), and those respondents not in poverty were more likely to use PHM than those in poverty (OR 1.24, 95% CI 0.97-1.58). Women were more likely to use PHM compared with men (OR 1.42, 95% CI 1.25-1.62), and respondents who reported a non-Hispanic ethnicity were more likely to use PHM than those reporting a Hispanic ethnicity (OR 1.4, 95% CI 1.14-1.71). Respondents living in the western (OR 1.52, 95% CI 1.3-1.77) and southern (OR 1.27, 95% CI 1.06-1.51) regions of the United States had an increased likelihood of PHM use compared with those living in the Midwest region. Frequency of anxiety was also a predictor of PHM use.

Individuals reporting anxiety on a weekly (OR 1.46, 95% CI 1.19-1.79), monthly (OR 1.44, 95% CI 1.17-1.77), or daily (OR 1.29, 95% CI 1.01-1.65) basis were more likely to use PHM than those reporting never having anxiety. Respondents reporting daily levels of moderate PA were more likely to use PHM than those who never engage in moderate PA (OR 1.44, 95% CI 1.21-1.72), and individuals who report experiencing barriers to accessing health care were more likely to use PHM than those who reported no barriers to accessing health care (OR 1.35, 95% CI 1.14-1.59). Finally, individuals who worked at organizations with more than 50 employees (OR 1.33, 95% CI 1.19-1.49) were somewhat more likely to use PHM than those who reported working at organizations with 50 employees or less.

Multiple Chronic Conditions Subgroup

PHM use for individuals reporting multiple chronic conditions was 14.7%. Among all factors listed in Tables 1-3, the highest proportional use of PHM in this group was among individuals who reported income greater than US $50,000 per year (25.22%, 1296/5138), had a college-level education (22.58%, 1643/7276), and had current employment (21.93%, 1093/4984). Those reporting being in poverty (5.24%, 131/2500), without college-level education (5.94%, 386/6495), without health insurance (7.48%, 79/1055), and those who were born outside the United States (7.54%, 122/1616) had the lowest proportional use of PHM. Education level and income were the 2 factors with the greatest difference in PHM use between levels. There was nearly a 17-point difference in PHM use between those who earned less than US $50 per year and those who earned more than US $50 per year and those with college-level education and those without college-level education.

Figure 4 illustrates the predictors of PHM use among individuals reporting multiple chronic conditions and is ordered by the magnitude of the no chronic condition. It indicates that college-level education was the strongest predictor of PHM use among individuals with multiple chronic conditions (OR 2.85, 95% CI 2.46-3.31), and individuals who reported family incomes greater than US $50 were more likely to use PHM than those earning less (OR 1.92, 95% CI 1.65-2.23). Similarly, individuals who reported not being in poverty were more likely to use PHM than those in poverty (OR 1.60, 95% CI 1.27-2.02). Respondents with health insurance were nearly twice as likely to use PHM than those without health insurance (OR 1.82, 95% CI 1.34-2.48), and individuals with a usual place of care had an increased likelihood of PHM use compared with those without a usual place of care (OR 1.87, 95% CI 1.28-2.72). Respondents who were born in the United States were more likely to use PHM than those who were born outside the United States (OR 1.56, 95% CI 1.23-1.97). Individuals who reported living in the west were more likely to use PHM than those in the Midwest (OR 1.47, 95% CI 1.26-1.71). Non-Hispanic respondents were more likely to use PHM than respondents who were Hispanic (OR 1.42, 95% CI 1.14-1.78), and white were more likely to use PHM than non-whites (OR 1.3, 95% CI 1.11-1.52). Individuals aged 18 to 40 years were more likely to use PHM than those older than 60 years (OR 1.54, 95% CI 1.25-1.88). Anxiety was also a predictor of PHM use. Individuals reporting daily (OR 1.45, 95% CI 1.13-1.86) or weekly anxiety (OR 1.33, 95% CI 1.06-1.66) were more likely to use PHM than individuals reporting never having anxiety. Respondents who reported never smoking were more likely to use PHM than those who reported currently smoking (OR 1.39, 95% CI 1.15-1.66). Individuals reporting current light alcohol use (OR 1.46 95% CI 1.28-1.67) or current heavy alcohol use (OR 1.42, 95% CI 1.09-1.83) were more likely to use PHM than nonalcoholic consumers. Respondents who reported working at organizations with more than 50 employees were more likely to use PHM (OR 1.26, 95% CI 1.13-1.4) than those who reported working at organizations with 50 or less employees, and individuals who reported being employed were more likely to use PHM than individuals who reported being unemployed (OR 1.27, 95% CI 1.09-1.48). Individuals who reported engaging in PA daily were more likely to use PHM than those reporting never engaging in PA (OR 1.21, 95% CI 1.0-1.47).

Comparisons Across Groups

For the logistic regressions, college education consistently had the largest ORs, with analyses from all subgroups reporting respondents with a college education being nearly 3 times as likely to report PHM use. Interestingly, women reporting no chronic conditions or 1 chronic condition were more likely to use PHM than men, and only among respondents reporting multiple chronic conditions was race found to be a significant predictor. Table 4 reports the ORs and CIs for all significant factors across the no chronic condition, single chronic condition, and multiple chronic conditions groups. There were few differences in significant predictors between the 3 models. On the basis of nonoverlapping CIs, the odds of using PHM were higher for those with no chronic conditions who reported no other barriers to accessing care than those reporting the same with 1 or more chronic conditions. Similarly, based on a slight overlap in CIs (0.05), the odds of using PHM among those with higher family incomes were higher for those with multiple chronic conditions compared with those with no chronic conditions.

Table 4. Comparison of patient health management (PHM) use by significant characteristics between respondents reporting no chronic conditions, 1 chronic condition, and multiple chronic conditions. Italics indicate nonsignificant findings.

Variable category and name	No chronic condition (n=22,929), OR[a] (95% CI)	One chronic condition (n=12,415), OR (95% CI)	Multiple chronic conditions (n=12,470), OR (95% CI)
Demographics			
Age in years			
18-40	1.3 0 (1.09 -1.57)[b]	1.58 (1.33 -1.87)[b]	1.64 (1.34-2.00)[b]
Sex			
Female	1.63 (1.49-1.79)	1.5 0 (1.34-1.67)	*1.04 (0.93-1.16)*
Race			
White	*1.06 (0.94-1.19)*	*1.08 (0.94-1.25)*	1.26 (1.11-1.46)
Ethnicity			
Not Hispanic	1.16 (1.02-1.34)	1.35 (1.11-1.64)	1.39 (1.11-1.75)
US born			
Yes	*1.04 (0.92-1.17)*	*1.01 (0.85-1.20)*	1.54 (1.23 -1.94)
Geography			
South	*1.11 (0.98-1.25)*	1.19 (1.03-1.38)	*1.03 (0.89-1.20)*
West	1.49 (1.32-1.69)	1.47 (1.26-1.72)	1.49 (1.27 -1.74)
Socioeconomic status			
Education			
College	2.65 (2.33-3.03)	2.80 (2.42-3.25)	2.81 (2.46 -3.20)
Family income			
US $50,000+	1.60 (1.43-1.79)	1.67 (1.46 -1.91)	2.03 (1.78 -2.31)
Poverty			
No	1.31 (1.11-1.55)	1.25 (1.01 -1.55)	1.59 (1.28 -1.99)
Number of employees			
51+	1.41 (1.29-1.54)	1.31 (1.18 -1.46)	1.32 (1.19-1.47)
Employed			
Yes	1.25 (1.11 -1.41)	1.16 (1.01 -1.34)	1.22 (1.07-1.40)
Insurance			
Yes	2.13 (1.79 -2.56)	1.74 (1.40-2.17)	1.78 (1.35 -2.36)
Housing			
Do not own	1.17 (1.05-1.29)	*1.10 (0.96-1.25)*	*1.03 (0.90-1.18)*
Food security			
Secure	1.51 (1.15-2.02)	1.51 (1.15-2.02)	*1.11 (0.89-1.39)*
Other barriers			
Yes	1.75 (1.53 -2.01)	1.31 (1.11-1.53)	1.22 (1.05 -1.41)
Health-related characteristics			
Health status			
Fair or poor	1.34 (1.01 -1.82)	*1.06 (0.86-1.30)*	*1.05 (0.90-1.21)*
Usual place of care			
Yes	1.82 (1.56-2.08)	1.59 (1.29 -1.97)	1.76 (1.25-2.54)
Alcohol			
Current light	1.37 (1.23-1.53)	1.54 (1.35-1.76)	1.36 (1.21-1.54)

Variable category and name	No chronic condition (n=22,929), OR[a] (95% CI)	One chronic condition (n=12,415), OR (95% CI)	Multiple chronic conditions (n=12,470), OR (95% CI)
Current heavy	1.51 (1.23-1.84)	1.56 (1.22-1.99)	1.35 (1.05 -1.72)
Smoking			
Never	1.38 (1.20-1.59)	1.6 0 (1.36-1.89)	1.36 (1.15 -1.62)
Depression			
Few times per year	1.14 (1.02-1.27)	*1.04 (0.90-1.19)*	*1.11 (0.96-1.28)*
Daily	1.5 0 (1.05 -2.11)	*1.13 (0.80-1.59)*	*1.15 (0.87-1.50)*
Anxiety			
Few times per year	1.31 (1.17-1.47)	*1.14 (0.99-1.31)*	*1.12 (0.97-1.30)*
Monthly	1.69 (1.45 -1.97)	1.36 (1.12 -1.66)	*1.06 (0.85-1.33)*
Weekly	1.62 (1.37 -1.91)	1.43 (1.17 -1.75)	1.29 (1.04 -1.59)
Daily	1.41 (1.14 -1.74)	1.29 (1.01 -1.65)	1.37 (1.09-1.73)
Moderate physical activity level			
Daily	1.45 (1.26-1.67)	1.42 (1.19-1.68)	1.25 (1.06-1.48)

[a] OR: odds ratio.

[b] The age range of 18-40 years had significantly higher odds than both other age categories.

Discussion

Principal Findings

Web-based interactions between patients and health organizations related to the access of health information are becoming a focused area of attention. Between 2009 and 2014, there was an increase in the use of secure email to communicate with HCPs [38]. The impact of patient access and use of health information has been thoroughly documented in the literature [39]. However, more detailed studies have shown that individuals who engaged with PHM were more likely to report an improved experience of care, improved outcomes, and improved health literacy. Despite increased attention to engaging patients and families in health decision making by providing access to electronic health information, there has been an overall low level of use of these tools by patients for PHM. Findings from this study support that assertion in that it found only 12.2% of American adults reported PHM use between 2009 and 2014, and the percentage of PHM use among this group has remained relatively stable over this period—only increasing slightly from less than 2% points from 2009 to 2014. This slight increase may be partly explained by the adoption of EHR systems that have the ability to engage patients. For example, the rate of adoption of systems with patient engagement functionality increased from 28% to 40% between 2009 and 2012 [40].

This study demonstrated that there are differences in PHM use across demographic, socioeconomic, and health-related individual-specific factors, and the proportions of persons reporting PHM use are different across groups with no chronic conditions versus groups with 1 or multiple chronic conditions. A greater proportion of individuals with chronic health conditions reported PHM use compared with those without chronic health conditions, and this finding is consistent with prior research [41]. The use of PHM by a greater proportion of individuals suffering from chronic conditions may be partly explained by clinical need. Individuals with chronic conditions may have complex treatment plans, they seek health care more frequently, and the management of chronic conditions typically requires medication management [42]. Thus, individuals accessing and managing care more frequently may result in greater use of scheduling appointments on the Web, requesting prescription refills on the Web, or communicating with HCPs on the Web.

Demographic Factors

Demographic factors clearly influenced PHM use across all 3 groups. Individuals who were younger, non-Hispanic, and who lived in the west reported the greatest levels of PHM use. Individuals who reported being born in the United States with multiple chronic conditions were more likely to use PHM compared with the other 2 models. Previous research has shown that white [7,25,28,29] and non-Hispanic individuals are more likely to use PHM than other racial or ethnic groups [23,25]. Our findings suggest a relationship between complexity of condition and PHM use. Previous research has also demonstrated that patients who are younger and non-Hispanic are more likely to be engaged in their health care. Our results confirm this previous research as related to individuals who reported a single chronic condition. Among the no chronic condition and single chronic condition groups, women were more likely to use PHM than men, and this finding is also consistent with previous research [7,23,25].

Socioeconomic Factors

Socioeconomic factors were the most predictive of PHM use across all 3 groups. Across all the groups, those with a college-level education were more than 2.65 times as likely to use PHM. Research has shown that there is a relationship between education level and health literacy and that health literacy increases a patient's engagement in health care decision making [28]. An individual's level of education is also associated with socioeconomic status. As socioeconomic status impacts an individual's health literacy, access to routine health care services, and also access to internet and computer technology, education level is a critical determinant in whether an individual emails an HCP, schedules an appointment on the Web, or requests a prescription refill on the Web [13,25]. There are clear disparities in PHM use based on education level. This issue deserves focused attention to ensure the disparities do not continue to widen over time. Similarly, insurance coverage and family income were among the strongest predictors of PHM use across the 3 groups, and these factors have been previously shown to be associated with patient engagement [7,43,44]. These findings are consistent with prior research regarding disparities in the use of patient-centered technology to connect patients to electronic health information [3,28]. Having a lower socioeconomic status may indicate the lack of internet access that would enable an individual to connect to their electronic health information [45]. Lower socioeconomic status may also indicate lack of employment and health insurance, thus an increased likelihood of encountering barriers to accessing health care and therefore electronic health information provided by health care organizations [13,25].

Health-Related Factors

Another predictor of PHM is having a usual place for receiving health care. Research has shown there is variation in EHR adoption nationally [46,47]. Different rates of EHR system adoption among both inpatient and ambulatory health care organizations may be impacting the PHM use. As PHM is associated with communicating with HCPs and using technology associated with health care organizations, having a usual source for receiving health care and therefore potentially having stronger relationships with HCPs on the surface would facilitate increased use of PHM [48]. Having a usual place of care also may indicate that these individuals routinely attend the same place for their clinical encounters and potentially obtain care from the same clinicians. This routine care from the same clinic and care team indicates that having an organization that one may consider a health care *home* produces an environment that promotes communication between clinicians and patients through technology.

Engaging in moderate PA was also associated with increased PHM use. Research has shown that there is a relationship between health status and patient engagement, and these findings suggest that individuals who are more active and therefore potentially healthier are more likely to engage in managing their health information through technology.

Interestingly, individuals with 1 chronic condition who reported current light alcohol use and daily moderate PA were more likely to use PHM than those who reported never consuming alcohol and never engaging in moderate PA. These findings can partly be explained by the association between health status and PHM use. Individuals who reported healthier lifestyle behavior such as PA, moderate alcohol consumption, and not smoking were also more likely to use PHM. Research has shown that patients who are less physically active are more likely to be disengaged in their health care [49]. Individuals who report never participating in moderate PA are less likely to use PHM than individuals who participate in moderate PA daily, which indicates a relationship between a general health-related lifestyle or physical ability and use of technology for accessing personal health information. Research has shown that there is a relationship between social determinants of health such as access to resources and an individual's level of PA [50].

Individuals who reported daily and weekly anxiety were more likely to use PHM than those who reported never having anxiety across all 3 subgroups. Previous research has found a relationship between anxiety and use of technology [51-53]. Research related to the Unified Theory of Acceptance and Use of Technology found that increased anxiety levels related to use of computers is negatively associated with behavioral intention to use technology [53]. Our findings suggest that there is an association between individual anxiety level and use of technology to email HCPs, schedule appointments on the Web, or request prescription refills on the Web. This may be partly explained by the relationship between chronic disease, stress, and anxiety [54]. Research has shown that *health anxiety* influences patient care and information-seeking behavior [55]. This finding suggests an association between individual concerns and individual behaviors surrounding electronic access and use of heath information. More anxious and more depressed patients are significantly more likely to use PHM. As anxiety and depression are often related, it makes some sense that these groups are seeking some reassurance from using PHM in any of its forms. It is important to reiterate that PHM consists of texting providers, refilling prescriptions, and scheduling appointments. It is possible that these individuals are attempting to consult with a provider to address their anxiety and depressive state.

Observations of Significant Factors Among Groups

This study demonstrated a small number of differences between the models for respondent subgroups. The odds of using PHM were higher among those without a chronic condition and those who do not experience barriers to accessing health care. This finding suggests that those individuals who do not experience challenges to accessing care and have chronic conditions may be using PHM to complement their typical access to HCPs or assist in managing their disease. Certain demographic factors also differed across models. Women without chronic conditions were more likely to use PHM than women with either 1 or multiple chronic conditions, and this finding is consistent with previous research [7,23,25,56].

Limitations

There are several limitations of this study. The findings are limited by the use of existing data collected through the NHIS survey for a purpose different from this study. The study depended upon the use of the standardized items included through self-reporting in the data collection process. Self-reported survey data have the potential to be biased by social desirability, leading to providing answers that the respondent perceives as more desirable. This could have led to an overestimate of the PHM use as compared with measurement of actual use. This study also relied on combining multiple years of NHIS data and adjusting sample weights to account for pooling data across years. This also could have led to an overestimation of PHM use based on certain factors. Our findings point to a direction for further, more sophisticated analysis of these data to further explore the findings. Another potential limitation of this study is the combined measure of PHM. The measure was constructed to reflect personal use of technology related to health care organizations, but the construction of the measure was limited to the 3 predefined items related to technology use. Furthermore, the questions asked only if the respondent had ever engaged in the activity, so responses did not reflect the intensity of use. This may lead to the underestimation of actual use, as respondents who reported engaging in the behavior based on only 1 incident are lumped with those who displayed the same behavior frequently. We conclude that the questions related to emailing HCPs, requesting prescription refills on the Web, and scheduling appointments are aggregable into a single PHM measure because of their high rates of co-occurrence in the dataset. There may also be additional measures that also reflect PHM that were not included in the dataset such as patient review of laboratory tests or visit summaries that were not included in the survey and have been reported previously. Leaving out these frequently reported behaviors could lead to an underestimate of use. These limitations of the dataset may separately lead to either over or underestimates of PHM use. We argue that this implies that our reported results are reasonable estimates but may have wider CIs than calculated. The NHIS did not collect information related to PHM in 2010, which may impact the overall proportional use of PHM across all years, but this was addressed by excluding this year from the analytic dataset.

Conclusions

The purpose of this research was to describe the overall utilization of PHM and compare individual characteristics associated with PHM in groups with no reported chronic conditions, with 1 chronic condition, and with 2 or more such conditions. The results indicate that the overall usage of PHM is not increasing along with the increased use of EHRs in the United States, even when clinical providers and hospitals are offering PHM features to patients. The overall use of PHM has increased slightly since 2009, but individuals reporting 1 or more chronic conditions used PHM at higher rates than individuals reporting no chronic conditions. The findings of this study also illustrated the disparities in PHM use across multiple factors, including economics and education in a nationally representative sample of individuals. These findings provide further evidence of the challenge associated with engaging patients through the use of electronic health information as the health care industry continues to evolve. Although health care organizations continue to adopt electronic modes of communication to facilitate interactions between patients and health care organizations, there are significant gaps related to the use of these tools for connecting consumers to health information. For each chronic condition category analyzed, demographic and socioeconomic factors appear to be driving PHM use. Research has demonstrated that patient-centered technologies are associated with improved clinical outcomes, patient experience, and health literacy. If action is not taken to address disparities in PHM use, individuals with lower socioeconomic status are at risk of seeing gaps in health disparities widen. In the short term, it is imperative that health care organizations develop initiatives aimed at promoting adoption of these tools by all individuals, regardless of socioeconomic status. Initiatives must be sensitive to health literacy, race, ethnicity, and other social determinants of health in their design if substantial progress in PHM use is to be achieved. In the long term, technologies that support PHM use must be designed to better meet the needs of patient populations. The current rate of use reflects the general lack of adoption of these tools, which can be partly explained by the lack of interest or need to access electronic health information on the Web. If PHM use is to increase over time, there needs to be better access to health care information across the continuum of care and more integration of tools and information related to personal fitness, diet, and lifestyle into the systems that support PHM use. Attention needs to be placed on developing technology solutions that meet the needs of all individuals regardless of educational achievement.

Future studies should investigate the relationship between PHM use and clinical outcomes across different chronic conditions, as well as the relationship between PHM use and usability of systems that support PHM-related functions. An additional level of analysis is warranted in which multiple significant predictors are clustered, rather than having 22 to 25 variables analyzed separately.

Previous research has indicated a relationship between the PHM and improved patient satisfaction, care outcomes, and knowledge. This work demonstrated that even though there was an overall low level of PHM use, there were clear disparities across demographic, socioeconomic, and health-related variables. The research showed that having a chronic condition is not the characteristic that best explains PHM use.

Acknowledgments

The authors wish to acknowledge the IHIS and the Minnesota Population Center at the University of Minnesota.

Conflicts of Interest

None declared.

References

1. Centers for Medicare & Medicaid Services (CMS), HHS. Medicare and Medicaid Programs; Electronic Health Record Incentive Program—Stage 3 and Modifications to Meaningful Use in 2015 Through 2017. Final rules with comment period. Fed Regist 2015 Oct 16;80(200):62761-62955 [FREE Full text] [Medline: 26477064]

2. Kohn LT, Corrigan JM, Donaldson MS. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: Committee on Quality of Health Care in America, Institute of Medicine; 2001.
3. Patel V, Barker W, Siminerio E. Disparities in Individuals? Access and Use of Health IT in 2013: ONC Data Brief No. 26. Washington, DC: Office of the National Coordinator for Health Information Technology; 2015. URL: https://www.healthit.gov/sites/default/files/briefs/oncdatabrief26june2015consumerhealthit.pdf [WebCite Cache ID 6slTbnUWN]
4. Hibbard JH, Baron R, Trisolini M. 2012. Patient Activation and Engagement for ACOs URL: https://acoregister.rti.org/index.cfm?fuseaction=dsp_web [accessed 2017-08-16] [WebCite Cache ID 6sla6ThWF]
5. Turvey CL, Zulman DM, Nazi KM, Wakefield BJ, Woods SS, Hogan TP, et al. Transfer of information from personal health records: a survey of veterans using My HealtheVet. Telemed J E Health 2012 Mar;18(2):109-114. [doi: 10.1089/tmj.2011.0109] [Medline: 22304439]
6. Sequist TD, Zaslavsky AM, Colditz GA, Ayanian JZ. Electronic patient messages to promote colorectal cancer screening: a randomized controlled trial. Arch Intern Med 2011 Apr 11;171(7):636-641 [FREE Full text] [doi: 10.1001/archinternmed.2010.467] [Medline: 21149743]
7. Tenforde M, Nowacki A, Jain A, Hickner J. The association between personal health record use and diabetes quality measures. J Gen Intern Med 2012 Apr;27(4):420-424 [FREE Full text] [doi: 10.1007/s11606-011-1889-0] [Medline: 22005937]
8. Simon GE, Ralston JD, Savarino J, Pabiniak C, Wentzel C, Operskalski BH. Randomized trial of depression follow-up care by online messaging. J Gen Intern Med 2011 Jul;26(7):698-704. [doi: 10.1007/s11606-011-1679-8] [Medline: 21384219]
9. Dorr D, Bonner LM, Cohen AN, Shoai RS, Perrin R, Chaney E, et al. Informatics systems to promote improved care for chronic illness: a literature review. J Am Med Inform Assoc 2007;14(2):156-163 [FREE Full text] [doi: 10.1197/jamia.M2255] [Medline: 17213491]
10. Turley M, Garrido T, Lowenthal A, Zhou YY. Association between personal health record enrollment and patient loyalty. Am J Manag Care 2012 Jul 01;18(7):e248-e253 [FREE Full text] [Medline: 22823553]
11. Finkelstein J, Knight A, Marinopoulos S, Gibbons MC, Berger Z, Aboumatar H, et al. Enabling patient-centered care through health information technology. Evid Rep Technol Assess (Full Rep) 2012 Jun(206):1-1531. [Medline: 24422882]
12. Gibbons MC, Wilson RF, Samal L, Lehman CU, Dickersin K, Lehmann HP, et al. Impact of consumer health informatics applications. Evid Rep Technol Assess (Full Rep) 2009 Oct(188):1-546. [Medline: 20629477]
13. Feinberg I, Greenberg D, Frijters J. 2015. Understanding Health Information Seeking Behaviors of Adults with Low Literacy, Numeracy, and Problem Solving Skills: Results from the 2012 US PIAAC Study URL: https://static1.squarespace.com/static/51bb74b8e4b0139570ddf020/t/54da7889e4b004762df45b52/1423603849970/Feinberg_Greenberg_Frijters_PIAAC.pdf [WebCite Cache ID 6sn6akrCf]
14. Jamal A, Khan SA, AlHumud A, Al-Duhyyim A, Alrashed M, Bin Shabr F, et al. Association of online health information-seeking behavior and self-care activities among type 2 diabetic patients in Saudi Arabia. J Med Internet Res 2015;17(8):e196 [FREE Full text] [doi: 10.2196/jmir.4312] [Medline: 26268425]
15. Ryu B, Kim N, Heo E, Yoo S, Lee K, Hwang H, et al. Impact of an electronic health record-integrated personal health record on patient participation in health care: development and randomized controlled trial of MyHealthKeeper. J Med Internet Res 2017 Dec 07;19(12):e401 [FREE Full text] [doi: 10.2196/jmir.8867] [Medline: 29217503]
16. Charles D, Gabriel M, Henry J. Electronic Capabilities for Patient Engagement among U.S. Non-Federal Acute Care Hospitals: ONC Data Brief No. 29. Washington DC: Office of the National Coordinator for Health Information Technology; 2015. URL: https://www.healthit.gov/sites/default/files/briefs/oncdatabrief29_patientengagement.pdf [WebCite Cache ID 6sn6o097z]
17. Patel V, Barker W, Siminerio E. Trends in Consumer Access and Use of Electronic Health Information: ONC Data Brief No. 30. Washington, DC: Office of the National Coordinator for Health Information Technology; 2015. URL: https://www.healthit.gov/sites/default/files/briefs/oncdatabrief30_accesstrends_.pdf [WebCite Cache ID 6slc3XTca]
18. Riippa I, Linna M, Rönkkö I, Krüger V. Use of an electronic patient portal among the chronically ill: an observational study. J Med Internet Res 2014 Dec 08;16(12):e275 [FREE Full text] [doi: 10.2196/jmir.3722] [Medline: 25488754]
19. Redelmeier DA, Kraus NC. Patterns in patient access and utilization of online medical records: analysis of MyChart. J Med Internet Res 2018 Feb 06;20(2):e43 [FREE Full text] [doi: 10.2196/jmir.8372] [Medline: 29410386]
20. Nøhr C, Parv L, Kink P, Cummings E, Almond H, Nørgaard JR, et al. Nationwide citizen access to their health data: analysing and comparing experiences in Denmark, Estonia and Australia. BMC Health Serv Res 2017 Aug 07;17(1):534 [FREE Full text] [doi: 10.1186/s12913-017-2482-y] [Medline: 28784173]
21. Gheorghiu B, Hagens S. Use and maturity of electronic patient portals. Stud Health Technol Inform 2017;234:136-141. [Medline: 28186030]
22. Westin A, Lemieux J, Subias S. Markle.org. Americans overwhelmingly believe electronic personal health records could improve their health URL: https://tinyurl.com/y8ewr9t4 [accessed 2017-08-17] [WebCite Cache ID 6sn6tA55y]
23. Wen KY, Kreps G, Zhu F, Miller S. Consumers' perceptions about and use of the internet for personal health records and health information exchange: analysis of the 2007 Health Information National Trends Survey. J Med Internet Res 2010;12(4):e73 [FREE Full text] [doi: 10.2196/jmir.1668] [Medline: 21169163]

24. Zhang Y, Lauche R, Sibbritt D, Olaniran B, Cook R, Adams J. Comparison of health information technology use between American adults with and without chronic health conditions: findings from the National Health Interview Survey 2012. J Med Internet Res 2017 Oct 05;19(10):e335 [FREE Full text] [doi: 10.2196/jmir.6989] [Medline: 28982644]
25. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, Lopez A, et al. The literacy divide: health literacy and the use of an internet-based patient portal in an integrated health system-results from the diabetes study of northern California (DISTANCE). J Health Commun 2010;15 Suppl 2:183-196 [FREE Full text] [doi: 10.1080/10810730.2010.499988] [Medline: 20845203]
26. Goel MS, Brown TL, Williams A, Cooper AJ, Hasnain-Wynia R, Baker DW. Patient reported barriers to enrolling in a patient portal. J Am Med Inform Assoc 2011 Dec;18 Suppl 1:i8-12 [FREE Full text] [doi: 10.1136/amiajnl-2011-000473] [Medline: 22071530]
27. Ralston JD, Carrell D, Reid R, Anderson M, Moran M, Hereford J. Patient web services integrated with a shared medical record: patient use and satisfaction corrected published erratum appears in J Am Med Inform Assoc 2008 Mar-Apr;15(2). J Am Med Inform Assoc 2007;14(6):265.
28. 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]
29. Ancker JS, Kern LM, Abramson E, Kaushal R. The Triangle Model for evaluating the effect of health information technology on healthcare quality and safety. J Am Med Inform Assoc 2012;19(1):61-65 [FREE Full text] [doi: 10.1136/amiajnl-2011-000385] [Medline: 21857023]
30. Greenberg AJ, Falisi AL, Finney Rutten LJ, Chou WS, Patel V, Moser RP, et al. Access to electronic personal health records among patients with multiple chronic conditions: a secondary data analysis. J Med Internet Res 2017 Jun 02;19(6):e188 [FREE Full text] [doi: 10.2196/jmir.7417] [Medline: 28576755]
31. Woods SS, Forsberg CW, Schwartz EC, Nazi KM, Hibbard JH, Houston TK, et al. The association of patient factors, digital access, and online behavior on sustained patient portal use: a prospective cohort of enrolled users. J Med Internet Res 2017 Oct 17;19(10):e345 [FREE Full text] [doi: 10.2196/jmir.7895] [Medline: 29042345]
32. Ward BW, Schiller JS, Goodman RA. Multiple chronic conditions among US adults: a 2012 update. Prev Chronic Dis 2014 Apr 17;11:E62 [FREE Full text] [doi: 10.5888/pcd11.130389] [Medline: 24742395]
33. IPUM Health Surveys. Minneapolis: University of Minnesota; 2012. URL: https://ihis.ipums.org/ihis/sda.shtml [accessed 2017-08-17] [WebCite Cache ID 6sn7ue2jH]
34. NHIS. CDC. 2012. National Health Interview Survey URL: https://www.cdc.gov/nchs/nhis/about_nhis.htm [accessed 2017-08-16] [WebCite Cache ID 6slcatcY7]
35. Johnson PJ, Blewett LA, Ruggles S, Davern ME, King ML. Four decades of population health data: the integrated health interview series as an epidemiologic resource. Epidemiology 2008 Nov;19(6):872-875 [FREE Full text] [doi: 10.1097/EDE.0b013e318187a7c5] [Medline: 18854709]
36. Iezzoni LI, Yu J, Wint AJ, Smeltzer SC, Ecker JL. Conditions causing disability and current pregnancy among US women with chronic physical disabilities. Med Care 2014 Jan;52(1):20-25 [FREE Full text] [doi: 10.1097/MLR.0000000000000015] [Medline: 24220686]
37. CDC. 2016. Variance estimation guidance, NHIS 2006-2015 URL: https://www.cdc.gov/nchs/data/nhis/2006var.pdf [accessed 2018-05-01] [WebCite Cache ID 6z5kqMn2v]
38. Sandefer RH, Khairat SS, Pieczkiewicz DS, Speedie SM. Using publicly available data to characterize consumers use of email to communicate with healthcare providers. Stud Health Technol Inform 2015;216:401-405. [Medline: 26262080]
39. Sandefer RH, Westra BL, Khairat SS, Pieczkiewicz DS, Speedie SM. Determinants of consumer eHealth information seeking behavior. AMIA Annu Symp Proc 2015;2015:1121-1129 [FREE Full text] [Medline: 26958251]
40. ONC Budget Performance Measures. Washington, DC: Office of the National Coordinator for Health Information Technology; 2016. URL: https://dashboard.healthit.gov/apps/onc-budget-performance-measures.php [WebCite Cache ID 6sn7xSgIh]
41. Agarwal R, Anderson C, Zarate J, Ward C. If we offer it, will they accept? Factors affecting patient use intentions of personal health records and secure messaging. J Med Internet Res 2013;15(2):e43 [FREE Full text] [doi: 10.2196/jmir.2243] [Medline: 23470453]
42. Holman H, Lorig K. Patient self-management: a key to effectiveness and efficiency in care of chronic disease. Public Health Rep 2004;119(3):239-243 [FREE Full text] [doi: 10.1016/j.phr.2004.04.002] [Medline: 15158102]
43. Wade-Vuturo AE, Mayberry LS, Osborn CY. Secure messaging and diabetes management: experiences and perspectives of patient portal users. J Am Med Inform Assoc 2013 May 1;20(3):519-525 [FREE Full text] [doi: 10.1136/amiajnl-2012-001253] [Medline: 23242764]
44. Bos-Touwen I, Schuurmans M, Monninkhof EM, Korpershoek Y, Spruit-Bentvelzen L, Ertugrul-van der Graaf I, et al. Patient and disease characteristics associated with activation for self-management in patients with diabetes, chronic obstructive pulmonary disease, chronic heart failure and chronic renal disease: a cross-sectional survey study. PLoS One 2015;10(5):e0126400 [FREE Full text] [doi: 10.1371/journal.pone.0126400] [Medline: 25950517]

45. Roblin DW, Houston TK, Allison JJ, Joski PJ, Becker ER. Disparities in use of a personal health record in a managed care organization. J Am Med Inform Assoc 2009;16(5):683-689 [FREE Full text] [doi: 10.1197/jamia.M3169] [Medline: 19567790]
46. King J, Furukawa MF, Buntin MB. Geographic variation in ambulatory electronic health record adoption: implications for underserved communities. Health Serv Res 2013 Dec;48(6 Pt 1):2037-2059 [FREE Full text] [doi: 10.1111/1475-6773.12078] [Medline: 23800087]
47. DesRoches CM, Campbell EG, Rao SR, Donelan K, Ferris TG, Jha A, et al. Electronic health records in ambulatory care--a national survey of physicians. N Engl J Med 2008 Jul 3;359(1):50-60. [doi: 10.1056/NEJMsa0802005] [Medline: 18565855]
48. Bhandari N, Shi Y, Jung K. Seeking health information online: does limited healthcare access matter? J Am Med Inform Assoc 2014;21(6):1113-1117 [FREE Full text] [doi: 10.1136/amiajnl-2013-002350] [Medline: 24948558]
49. Carroll JK, Antognoli E, Flocke SA. Evaluation of physical activity counseling in primary care using direct observation of the 5As. Ann Fam Med 2011;9(5):416-422 [FREE Full text] [doi: 10.1370/afm.1299] [Medline: 21911760]
50. Adler NE, Newman K. Socioeconomic disparities in health: pathways and policies. Health Aff (Millwood) 2002;21(2):60-76 [FREE Full text] [Medline: 11900187]
51. Diefenbach MA, Miller SM, Daly MB. Specific worry about breast cancer predicts mammography use in women at risk for breast and ovarian cancer. Health Psychol 1999 Sep;18(5):532-536. [Medline: 10519469]
52. Beaudry A, Pinsonneault A. The other side of acceptance: studying the direct and indirect effects of emotions on information technology use. MIS Q 2010 Dec;34(4):689-710 [FREE Full text] [doi: 10.2307/25750701]
53. Attuquayefio S, Addo H. Eujournal. Review of studies with UTAUT as conceptual framework URL: https://eujournal.org/index.php/esj/article/view/3020 [WebCite Cache ID 6sn8GNPKL]
54. Garrido M, Hash-Converse JM, Leventhal H, Leventhal E. Stress and chronic disease management. In: The Handbook of Stress Science: Biology, Psychology, and Health. New York City: Springer Publishing Company; 2011:488-500.
55. Eastin MS, Guinsler NM. Worried and wired: effects of health anxiety on information-seeking and health care utilization behaviors. Cyberpsychol Behav 2006 Aug;9(4):494-498. [doi: 10.1089/cpb.2006.9.494] [Medline: 16901253]
56. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, López A, et al. Social disparities in internet patient portal use in diabetes: evidence that the digital divide extends beyond access. J Am Med Inform Assoc 2011 May 1;18(3):318-321 [FREE Full text] [doi: 10.1136/jamia.2010.006015] [Medline: 21262921]

Abbreviations

CMS: Centers for Medicare and Medicaid Services
EHR: electronic health record
HCP: health care provider
IHIS: Integrated Health Interview Series
NHIS: National Health Interview Survey
OR: odds ratio
PA: physical activity
PHM: personal health management

Edited by G Eysenbach; submitted 19.09.17; peer-reviewed by C Kruse, M Duplaga, C Turvey; comments to author 09.11.17; revised version received 01.05.18; accepted 18.06.18; published 18.10.18.

Please cite as:
Sandefer RH, Westra BL, Khairat SS, Pieczkiewicz DS, Speedie SM
Assessment of Personal Health Care Management and Chronic Disease Prevalence: Comparative Analysis of Demographic, Socioeconomic, and Health-Related Variables
J Med Internet Res 2018;20(10):e276
URL: http://www.jmir.org/2018/10/e276/
doi:10.2196/jmir.8784
PMID:30341046

Original Paper

Preferences for Health Information Technologies Among US Adults: Analysis of the Health Information National Trends Survey

Onur Asan[1], PhD; Farion Cooper II[2], BA; Sneha Nagavally[2], MS; Rebekah J Walker[2], PhD; Joni S Williams[2], MD, MPH; Mukoso N Ozieh[2], MD, MSc; Leonard E Egede[2], MD, MS

[1]School of Systems and Enterprises, Stevens Institute of Technology, Hoboken, NJ, United States

[2]Center for Advancing Population Science (CAPS), Medical College of Wisconsin, Milwaukee, WI, United States

Corresponding Author:
Onur Asan, PhD
School of Systems and Enterprises
Stevens Institute of Technology
1 Castle Point Terrace
Hoboken, NJ, 07030
United States
Phone: 1 2012165514
Email: oasan@stevens.edu

Abstract

Background: Emerging health technologies are increasingly being used in health care for communication, data collection, patient monitoring, education, and facilitating adherence to chronic disease management. However, there is a lack of studies on differences in the preference for using information exchange technologies between patients with chronic and nonchronic diseases and factors affecting these differences.

Objective: The purpose of this paper is to understand the preferences and use of information technology for information exchange among a nationally representative sample of adults with and without 3 chronic disease conditions (ie, cardiovascular disease [CVD], diabetes, and hypertension) and to assess whether these preferences differ according to varying demographic variables.

Methods: We utilized data from the 2012 and 2014 iteration of the Health Information National Trends Survey (N=7307). We used multiple logistic regressions, adjusting for relevant demographic covariates, to identify the independent factors associated with lower odds of using health information technology (HIT), thus, identifying targets for awareness. Analyses were weighted for the US population and adjusted for the sociodemographic variables of age, gender, race, and US census region.

Results: Of 7307 participants, 3529 reported CVD, diabetes, or hypertension. In the unadjusted models, individuals with diabetes, CVD, or hypertension were more likely to report using email to exchange medical information with their provider and less likely to not use any of the technology in health information exchange, as well as more likely to say it was not important for them to access personal medical information electronically. In the unadjusted model, additional significant odds ratio (OR) values were observed. However, after adjustment, most relationships regarding the use and interest in exchanging information with the provider were no longer significant. In the adjusted model, individuals with CVD, diabetes, or hypertension were more likely to access Web-based personal health information through a website or app. Furthermore, we assessed adjusted ORs for demographic variables. Those aged >65 years and Hispanic people were more likely to report no use of email to exchange medical information with their provider. Minorities (Hispanic, non-Hispanic black, and Asian people) were less likely to indicate they had no interest in exchanging general health tips with a provider electronically.

Conclusions: The analysis did not show any significant association among those with comorbidities and their proclivity toward health information, possibly implying that HIT-related interventions, particularly design of information technologies, should focus more on demographic factors, including race, age, and region, than on comorbidities or chronic disease status to increase the likelihood of use. Future research is needed to understand and explore more patient-friendly use and design of information technologies, which can be utilized by diverse age, race, and education or health literacy groups efficiently to further bridge the patient-provider communication gap.

(J Med Internet Res 2018;20(10):e277) doi:10.2196/jmir.9436

KEYWORDS

chronic conditions; eHealth; health information exchange; health information technology; internet; mHealth; mobile phone

Introduction

The burden of chronic diseases continues to grow, and they remain the most common cause of death and disability worldwide [1]. More than half of the adults in the United States have at least one chronic condition (eg, hypertension, coronary heart disease, stroke, diabetes, and cancer) [2,3]. Cardiovascular disease (CVD) and diabetes are two common chronic conditions that share common risk factors including age, genetic factors, obesity, poor nutrition, impaired glucose tolerance, and physical inactivity [4]. Moreover, diabetes is a well-known risk factor for CVD as it can increase atherosclerosis and cause elevated blood pressure and cholesterol levels in many individuals [4]. In addition, hypertension is a well-established risk factor for CVD, and studies have shown a relationship between hypertension and both vascular and structural cardiac remodeling [5]. Furthermore, inflammatory mediators such as damage-associated molecular patterns have been associated with CVD, hypertension, and diabetes [6]. Immune-mediated and adaptive immune responses have been implicated in hypertension and other vascular diseases as well [7]. Death from CVD rose by >40% between 1990 and 2013 as a result of factors such as population increase, aging, and epidemiological changes in chronic diseases [8]. Thus, it is critical for the population with chronic diseases to maintain effective disease management to achieve an improved quality of life and health outcomes [9,10].

The need for continuous monitoring, long-term planning, and frequent interaction of patients with chronic diseases with their providers might be addressed with ongoing technological advancements [1]. Emerging health technologies such as remote patient monitoring systems, mobile phone intervention programs, electronic health databases, smartphone apps, instant messaging, video calling, and patient portals are increasingly being used in health care for communication, data collection, patient monitoring, education, and facilitating adherence to chronic disease management [11-28]. It has been suggested that these technologies can assist in reducing both the burden and cost of CVD around the world [29-31].

Specifically, the use of Web-based and mobile health (mHealth) intervention programs has led to a few suggestions for the technological intervention to improve health information exchange (HIE). In fact, it has been shown that information exchange via various portals and mobile intervention has led to an increased adherence to healthy lifestyle programs, treatment regimens, and disease prevention programs [11-13,15,32]. However, the literature on this topic is very general and seldom concentrates on different chronic diseases and conditions. Literature that does converge on specific diseases like diabetes and chronic kidney disease only concentrate on one type of electronic health (eHealth) technology at a time [32,33]. The evidence that concentrates on the broad use of technology in addressing CVD specifically is limited [34]. There is also a lack of studies on differences in the preference for using information exchange technologies between patients with chronic and nonchronic diseases and factors affecting these differences. From a design and intervention development perspective, it would be critical to understand user differences and their relationship with demographic factors to augment the uptake of the newly developed information technologies.

The purpose of this paper is to understand the preferences and use of technology for information exchange among a nationally representative sample of adults with and without 3 chronic disease conditions (ie, CVD, diabetes, and hypertension) and to assess whether these preferences differ according to varying demographic variables.

Methods

Sample Population

Data for this study were derived from the Health Information National Trends Survey (HINTS). HINTS is a large-scale, household interview survey of US noninstitutionalized adults aged ≥18 years. HINTS gathers information from the general population to investigate trends in the utilization of health communication systems between providers and patients, specifically pertaining to access and usage [35]. We utilized data from HINTS 4 Cycles 2 (2012) and 4 (2014), which were obtained using the single mode mailing service and the Next Birthday Method for respondent selection. To perform a consistent selection of respondents across all households, we used the Next Birthday Method; in this technique, adults who have a next birthday coming up are requested to finish the survey for every family unit.

The sampling frame came from a collection of databases used by the Marketing Systems Group to obtain a random sample of addresses. This was then grouped into 3 specific sampling strata: (1) high concentrations of minority populations; (2) low concentrations of minority populations; and (3) counties in Central Appalachia (regardless of minority populations) [35]. A total of 12,055 households received the 4-part mailed questionnaire in Cycle 2, and 13,996 households received the 4-part mailed questionnaires in Cycle 4. The response rate was 39.97% (4818/12,055) for Cycle 2 and 34.44% (4820/13,996) for Cycle 4 [36,37]. Additional information about data collection and methodologies can be found in the corresponding methodology reports for HINTS 4 Cycles 2 and 4. The sampling methodology allowed for weighting of the sample to provide population estimates [36,37]. Further details on survey design and sampling strategies of the overall HINTS mechanism were published in a previous work [38].

We identified individuals with CVD, diabetes, or hypertension through self-reported diagnosis via the questions “Has a doctor or other health professional ever told you that you have diabetes or high blood sugar,” “Has a doctor or other health professional ever told you that you have high blood pressure or hypertension?” and “Has a doctor or other health professional ever told you that you have a heart condition such as heart attack, angina, or congestive heart failure?" Individuals included in the sample population answered yes to any of the diagnoses. *"Missing data"* or "Inapplicable" response type for these three questions was considered to be a missing value for having a chronic condition. These 2 versions of HINTS resulted in a sample of 7307 total individuals who answered relevant questions and for whom demographic data were collected, and this was the sample size used for this analysis [36,37].

Measures

The main outcome variables were questions asking about exchanging information with providers and accessing medical records. The questions that were included in the analysis and the responses dichotomized have been detailed in Multimedia Appendix 1.

We modeled negative responses using multiple logistic regressions adjusting for relevant covariates to identify independent factors associated with lowers odds of using health information technology (HIT) and, therefore, identify targets for awareness. Analyses were weighted for the US population and adjusted for the sociodemographic variables of age, gender, race, and US census region. Covariates were categorized as age (18-34 years, 35-44 years, 45-64 years, and 65-110 years); gender (female, male); race (Hispanic, non-Hispanic white, non-Hispanic black or African American, non-Hispanic Asian, and non-Hispanic other [non-Hispanic American Indian or Alaska Native or non-Hispanic, Native Hawaiian or other Pacific Islander or non-Hispanic multiple races mentioned]); and region (Northeast, Midwest, South, and West).

Statistical Analysis

First, basic demographic statistics were conducted using demographic attributes: age ("18-34," "35-44," "45-64," "≥65" years), gender ("Male," "Female"), race ("Hispanic," "non-Hispanic white," "non-Hispanic black or African American," "non-Hispanic Asian," "non-Hispanic other"), and census region ("Northeast," "Midwest," "South," "West"). We calculated weighted population frequencies and percentiles. Second, we analyzed each question with an unadjusted logistic regression model, followed by an adjusted logistic regression model. To model for dichotomous outcome variables, we used logit model. In a logit model, log odds of a dependent variable are modeled as a linear combination of independent variables. To implement the logit model, we used generalized linear model function and linked the binomial family to logistic regression to develop a logistic model for the analyses. Models were adjusted for age, race, gender, and region. All analyses were conducted using R statistical tool, with a $P<.05$ considered statistically significant. To analyze the weighted sample, we implemented the R Survey package with type "JKn" to include the weight samples across the dataset. Furthermore, we used sample weights from the survey data to analyze weighted population estimates and replicate weights to compute SE of estimates using the jackknife replication method.

Results

In all, 7307 participants (weighted population of 230,993,888) answered the questions used in this analysis. Within this sample, 3529 participants (weighted population of 90,748,995) reported CVD, diabetes, or hypertension. In addition, 3.33% (243/7307) of observations were with missing data. Table 1 shows the demographics of the population. Of those with CVD, diabetes, or hypertension, 42.35% (1482/3529) were aged 45-64 years and 31.06% (1094/3529) were aged ≥65 years. Furthermore, non-Hispanic white people represented 60.79% (2142/3529) of the population, and the Southern region of the United States represented 39.25% (1383/3529) of the population.

Table 2 displays the unadjusted and adjusted odds ratios (ORs) for HIT use and interest in exchanging information with a provider. In the unadjusted models, individuals with diabetes, CVD, or hypertension were more likely to report using email to exchange medical information with their provider (OR 1.431; 95% CI 1.113-1.838) and less likely to not use any of the technology in HIE (OR 0.778; 95% CI 0.618-0.979). Additionally, those with diabetes, CVD, or hypertension were less likely to report that health apps never led to asking new questions of their provider (OR 0.526; 95% CI 0.331-0.838) and were also more likely to report having no interest in exchanging electronic appointment reminders (OR 1.818; 95% CI 1.388-2.380), in exchanging general health tips with health care provider electronically (OR 1.283; 95% CI 1.004-1.639), in exchanging medication reminders with health care provider electronically (OR 1.440; 95% CI 1.135-1.827), in exchanging diagnostic information with health care provider electronically (OR 1.269; 95% CI 1.011-1.592), or in exchanging symptoms with health care provider electronically (OR 1.263; 95% CI 1.004-1.588). However, after adjustment, all relationships regarding the use of and interest in exchanging information with the provider were no longer significant.

Table 3 displays the unadjusted and adjusted ORs for interest in accessing personal medical information. In the unadjusted models, individuals with CVD, diabetes, or hypertension were more likely to say that it was not important for them to access personal medical information electronically (OR 1.496; 95% CI 1.142-1.959); however, in the adjusted models, these individuals were more likely to access Web-based personal health information through a website or app (OR 1.877; 95% CI 1.210-2.912).

Because a number of relationships lost significance after adjustment for demographic covariates, Multimedia Appendix 2 displays the adjusted ORs for demographic variables included in models presented in Tables 2 and 3 that showed significant unadjusted differences by chronic disease diagnosis to determine which demographic variables explained the relationship. Models in Multimedia Appendix 2 were adjusted by demographic covariates age, gender, race or ethnicity, and census region and the primary variable (presence of diabetes, CVD, or hypertension). Individuals aged >65 years (OR 2.32; 95% CI 1.55-3.49) and Hispanic people (OR 1.95; 95% CI 1.26-3.01) were more likely to report no use of email to exchange medical information with their provider. Conversely, non-Hispanic black people were less likely to report that they never used an app that led to questions for their provider (OR 0.32; 95% CI 0.16-0.63), whereas individuals living in the Western region of the country were more likely to report never using an app that led to questions or a second opinion (OR 2.18; 95% CI

1.08-4.41). Minorities were less likely to indicate that they had no interest in exchanging general health tips with a provider electronically (Hispanic people [OR 0.55; 95% CI 0.34-0.88], non-Hispanic black people [OR 0.36; 95% CI 0.23-0.56], and Asian people [OR 0.33; 95% CI 0.16-0.70]). In addition, non-Hispanic black people were less likely to report having no interest in exchanging medication reminders with providers (OR 0.62; 95% CI 0.41-0.95). Individuals aged 45-64 and >65 years were more likely to report no interest in exchanging a variety of medical information with their providers and more likely to say that it was not important to access their own information electronically. However, individuals aged >65 years were less likely to say that they never accessed personal medical information electronically in the past 12 months (OR 0.25; 95% CI 0.13-0.47). Finally, women were more likely to report no interest in exchanging diagnostic information with their provider (OR 1.32; 95% CI 1.03-1.71).

Table 1. Weighted population demographic characteristics: sample n=7307, weighted population N=230,993,888.

Demographics	Weighted population, n (%)		*P* value
	Population with diabetes, CVD[a], or hypertension (n=3529)	Population without diabetes, CVD, or hypertension (n=3535)	
Gender			.59
Male	1683 (47.76)	1672 (47.34)	
Female	1845 (50.19)	1863 (51.53)	
Age (years)			<.001[b]
18-34	351 (9.83)	1512 (42.80)	
35-44	458 (13.19)	671 (19.02)	
45-64	1482 (42.35)	996 (28.19)	
>65	1094 (31.06)	267 (7.63)	
Race			.12
Hispanic	413 (11.74)	530 (14.98)	
Non-Hispanic white	2142 (60.79)	2177 (61.63)	
Non-Hispanic black or African American	399 (11.35)	335 (9.48)	
Non-Hispanic Asian	92 (2.61)	198 (5.62)	
Non-Hispanic other	81 (2.34)	70 (1.96)	
Region			.22
Northeast	620 (17.58)	664 (18.86)	
Midwest	780 (22.16)	752 (21.28)	
South	1383 (39.25)	1260 (35.65)	
West	741 (21.00)	856 (24.22)	

[a] CVD: cardiovascular disease.

[b] Statistically significant relationships at *P*<.05, with the presence of diabetes, cardiovascular disease, or hypertension being the primary indicator.

Table 2. Unadjusted and adjusted odds ratios (ORs) for the use of and interest in exchanging information with a provider among individuals with the presence of a chronic condition.

Odds[a]	Unadjusted OR (95% CI)	Adjusted OR (95% CI)
Odds of reporting use		
In the past 12 months, used any of following to exchange info with health care professional?		
Email	1.431 (1.113-1.838)[b]	1.128 (0.865-1.470)
Short message service text message	1.039 (0.688-1.571)	0.955 (0.594-1.536)
App	0.983 (0.602-1.604)	0.734 (0.426-1.262)
Video	1.922 (0.560-6.599)	1.379 (0.310-6.134)
Social media	0.936 (0.499-1.756)	0.659 (0.325-1.337)
Fax	0.874 (0.553-1.379)	0.811 (0.503-1.308)
None	0.778 (0.618-0.979)[b]	0.963 (0.753-1.233)
Odds of responding "no" or "never"		
Have apps on smartphone or tablet related to health led to asking doctor new questions or getting second opinion from another doctor?	0.526 (0.331-0.838)[b]	0.591 (0.341-1.022)
How interested in exchanging appointment reminders with health care provider electronically?	1.818 (1.388-2.380)[b]	1.072 (0.723-1.588)
How interested in exchanging general health tips with health care provider electronically?	1.283 (1.004-1.639)[b]	1.037 (0.770-1.397)
How interested in exchanging medication reminders with health care provider electronically?	1.440 (1.135-1.827)[b]	0.975 (0.718-1.324)
How interested in exchanging lab or test results with health care provider electronically?	1.215 (0.951-1.552)	0.938 (0.683-1.289)
How interested in exchanging diagnostic information with health care provider electronically?	1.269 (1.011-1.592)[b]	0.996 (0.757-1.312)
How interested in exchanging vital signs with health care provider electronically?	1.259 (0.986-1.607)	1.022 (0.754-1.386)
How interested in exchanging lifestyle behaviors with health care provider electronically?	1.187 (0.942-1.496)	0.882 (0.669-1.163)
How interested in exchanging symptoms with health care provider electronically?	1.263 (1.004-1.588)[b]	0.832 (0.621-1.116)

[a] The model was adjusted by demographic covariates: age, gender, race or ethnicity, and census region.

[b] Statistically significant relationships at *P*<.05, with the presence of diabetes, cardiovascular disease, or hypertension being the primary indicator.

Table 3. Unadjusted and adjusted odds ratios (ORs) for use of and interest in accessing personal medical information among individuals with the presence of a chronic condition.

Odds[a]	Unadjusted OR (95% CI)	Adjusted OR (95% CI)
Odds of responding not important or not confident or no or more times		
How important is it for doctors to share your medical information with each other electronically?	0.960 (0.680-1.355)	1.282 (0.790-2.080)
How important would it be for you to access personal medical information electronically?	1.496 (1.142-1.959)[b]	1.089 (0.761-1.557)
How confident are you that safeguards are present to protect your medical records from being seen by people who are not permitted to view them?	0.817 (0.638-1.048)	0.902 (0.670-1.213)
Have you been offered access to personal health info through a website or app by your health care provider?	1.043 (0.837-1.301)	1.041 (0.817-1.324)
Odds of increasing number of times		
In the past 12 months, how many times did you access personal health information through a website or app?	1.251 (0.824-1.898)	1.877 (1.210-2.912)[b]

[a] The model was adjusted by demographic covariates: age, gender, race or ethnicity, and census region.

[b] Statistically significant relationships at *P*<.05, with the presence of diabetes, cardiovascular disease, or hypertension being the primary indicator.

Discussion

Principal Findings

In this study, we examined and compared patients' attitudes toward the use of health information technologies for HIE between providers and patients with and without the chronic (diabetes, CVD, or hypertension) conditions. The analyses showed that while unadjusted differences existed in responses to the use of HIT based on the existence of chronic disease, the inclusion of the demographic factors of gender, race, age, and region of residence removed this statistical significance and may explain the differences. This suggests that demographic differences are more important than differences due to presence or absence of a chronic disease. This is particularly important as interventions are developed to increase the use of HIT across different population groups in the United States. Specifically, our findings suggest that future interventions should target the unique needs of the elderly and ethnic minorities regardless of their disease status.

Our findings show that older individuals, regardless of the presence of a chronic condition, indicated less interest in using HIT to communicate or exchange information with their provider. Previous studies have also reported similar findings. One study showed the rate of HIT use among those aged ≥65 years to be less than that among younger patients [39]. Another study also showed that older adults were less likely than younger adults to value the importance of Patient Health Records [33]. There might be several reasons behind these findings, and they include the following: (1) greater ease of use and comfort with technology among younger adults [40]; (2) poor usability, availability, and accessibility of HIT functions tailored to the elderly [41-43]; (3) lower health literacy in the elderly [44,45]; (4) lower educational attainment and income in the elderly [3,46]; and (5) unique challenges from having chronic diseases that are more prevalent in the elderly and that impede the use of technology (eg, arthritis, vision impairment) [47]. However, the potential of technology for better communication and disease management is clear [48]; therefore, given the burden of chronic disease in the elderly, it is imperative to develop more user-friendly interfaces to facilitate better use of HIT for this age group. In addition, the results showed that adults aged 65 years with chronic disease reported no interest in exchanging medical information through the Web with their doctors, although they stated that they accessed health info electronically in the past 12 months. This might show that this group would like to access information on the Web to get informed, but still prefer other type of communication channels to discuss with their providers, such as in person or via phone over electronic information exchange. Finally, a recent study has shown increased use of mHealth technologies by the elderly group [49]; likewise, this study also signals that if these technologies are designed to be more patient friendly and if they address cognitive load as well as understand the needs of elderly patients, we would witness increased use and more eHealth information exchange among this group too.

In addition, we found that minorities showed more interest in using HIT, particularly for general health tips and medication reminders. In addition, non-Hispanic black people were less likely to report that they never used an app that led to questions for their provider. Some previous studies have shown that minorities are less likely to seek out Web-based health information [48], but our results show that this trend may be changing given the advancement in technology and rate of possession of mobile phones among minority populations. A recent Pew Internet & American Life Project's 2013 Health Online study has also shown that minorities (Hispanic and non-Hispanic black people) reported using their mobile phones to access health information, especially information related to pregnancy and weight, more than non-Hispanic white people [50,51]. It is also interesting to see that given the increased use of smartphone among minorities to access health information through app or Web, they are not more likely to report or exchange information with their providers. This changing trend suggests that future interventions for ethnic minorities should take greater advantage of technology, especially eHealth and mobile technology, which can also address and solve some disparity problems in the long run.

Some past studies have shown mixed results on gender differences in the eHealth use and perception of HIE [52-54]. In our study, the only question indicating a difference by gender suggests that men may be more interested in using HIT to discuss diagnostic information with their providers. Finally, individuals with chronic conditions were more likely to access Web-based personal health information through a website or app for the last 12 months. One reason for this finding might be that individuals with no chronic condition did not need or have any issue requiring them to access their health records compared with individuals with chronic conditions.

Limitations

In this analysis, we used a large, population-based sample providing generalizable results and investigated a variety of HIT options; however, there are some limitations worth noting. First, the data were cross-sectional and, therefore, cannot offer information on causality. Second, the response rate for HINTS was 21%-35%, and therefore, it may have more selection bias than other national surveys. Local or regional studies that provide more detailed data with a higher response rate should be conducted to validate these findings. Finally, there are several possible confounders that may further explain the relationships noted that were not included in the dataset.

Conclusions

This study compares the HIT use for information exchange by individuals with and without chronic conditions (diabetes, CVD, or hypertension) and potential factors that influence HIT use. The findings did not show any significant association among those with comorbidities and their proclivity toward health information. This study suggests that HIT-related interventions, particularly the design of mHealth technologies, should focus more on demographic factors, including race, age, and region, than on chronic disease status or comorbidities to increase the likelihood of use. Future research is needed to understand and explore more patient-friendly use and design of mHealth apps, which can be utilized by various age, race, and education or health literacy groups efficiently to further bridge the communication gap between patients and their provider. A more qualitative exploration would be beneficial to identify why certain groups do and do not use HIT for HIE purposes.

Acknowledgments

This study was supported by the National Institute of Diabetes and Digestive and Kidney Diseases (Grant K24DK093699; Principal Investigator: LEE, MD, MS). This award was also funded by the Research and Education Program Fund, which is a component of the Advancing a Healthier Wisconsin endowment at the Medical College of Wisconsin (Principal Investigator: OA, PhD).

Authors' Contributions

OA, JSW, RJW, and LEE designed the study. SN acquired and analyzed the data. SN, RJW, and LEE developed the analyses, and OA, JSW, SN, RJW, MNO, and LEE interpreted the data. OA, JSW, RJW, MNO, and LEE critically revised the manuscript for important intellectual content.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Survey questions used in the analysis.

[PDF File (Adobe PDF File), 64KB - jmir_v20i10e277_app1.pdf]

Multimedia Appendix 2

Adjusted odds ratios for selected questions with demographics information.

[PDF File (Adobe PDF File), 46KB - jmir_v20i10e277_app2.pdf]

References

1. 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;17(2):e52 [FREE Full text] [doi: 10.2196/jmir.3951] [Medline: 25803266]
2. 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;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]
3. Zhang Y, Lauche R, Sibbritt D, Olaniran B, Cook R, Adams J. Comparison of Health Information Technology Use Between American Adults With and Without Chronic Health Conditions: Findings From The National Health Interview Survey 2012. J Med Internet Res 2017 Oct 05;19(10):e335 [FREE Full text] [doi: 10.2196/jmir.6989] [Medline: 28982644]
4. Tong B, Stevenson C. Comorbidity of cardiovascular disease, diabetes and chronic kidney disease in Australia. Cardiovascular Disease Series 2007;28:37 [FREE Full text]
5. Koene RJ, Prizment AE, Blaes A, Konety SH. Shared Risk Factors in Cardiovascular Disease and Cancer. Circulation 2016 Mar 15;133(11):1104-1114 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.115.020406] [Medline: 26976915]
6. De Miguel C, Rudemiller NP, Abais JM, Mattson DL. Inflammation and hypertension: new understandings and potential therapeutic targets. Curr Hypertens Rep 2015 Jan;17(1):507 [FREE Full text] [doi: 10.1007/s11906-014-0507-z] [Medline: 25432899]
7. Harrison D, Guzik T, Lob H. Inflammation, immunity, and hypertension. Hypertension 2011;57(2):132-140 [FREE Full text]
8. Roth GA, Forouzanfar MH, Moran AE, Barber R, Nguyen G, Feigin VL, et al. Demographic and epidemiologic drivers of global cardiovascular mortality. N Engl J Med 2015 Apr 02;372(14):1333-1341 [FREE Full text] [doi: 10.1056/NEJMoa1406656] [Medline: 25830423]
9. Duff OM, Walsh DM, Furlong BA, O'Connor NE, Moran KA, Woods CB. Behavior Change Techniques in Physical Activity eHealth Interventions for People With Cardiovascular Disease: Systematic Review. J Med Internet Res 2017 Aug 02;19(8):e281 [FREE Full text] [doi: 10.2196/jmir.7782] [Medline: 28768610]
10. Whitehead L, Seaton P. The Effectiveness of Self-Management Mobile Phone and Tablet Apps in Long-term Condition Management: A Systematic Review. J Med Internet Res 2016;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
11. PausJenssen AM, Spooner BA, Wilson MP, Wilson TW. Cardiovascular risk reduction via telehealth: a feasibility study. Can J Cardiol 2008 Jan;24(1):57-60 [FREE Full text] [Medline: 18209771]

12. Purcell R, McInnes S, Halcomb EJ. Telemonitoring can assist in managing cardiovascular disease in primary care: a systematic review of systematic reviews. BMC Fam Pract 2014 Mar 07;15:43 [FREE Full text] [doi: 10.1186/1471-2296-15-43] [Medline: 24606887]
13. Brewer LC, Kaihoi B, Zarling KK, Squires RW, Thomas R, Kopecky S. The use of virtual world-based cardiac rehabilitation to encourage healthy lifestyle choices among cardiac patients: intervention development and pilot study protocol. JMIR Res Protoc 2015;4(2):e39 [FREE Full text] [doi: 10.2196/resprot.4285] [Medline: 25857331]
14. Posadzki P, Mastellos N, Ryan R, Gunn LH, Felix LM, Pappas Y, et al. Automated telephone communication systems for preventive healthcare and management of long-term conditions. Cochrane Database Syst Rev 2016 Dec 14;12:CD009921 [FREE Full text] [doi: 10.1002/14651858.CD009921.pub2] [Medline: 27960229]
15. Coughlin SS, Prochaska JJ, Williams LB, Besenyi GM, Heboyan V, Goggans DS, et al. Patient web portals, disease management, and primary prevention. Risk Manag Healthc Policy 2017;10:33-40 [FREE Full text] [doi: 10.2147/RMHP.S130431] [Medline: 28435342]
16. Shaw RJ, Ferranti J. Patient-provider internet portals—patient outcomes and use. Comput Inform Nurs 2011 Dec;29(12):714-8; quiz 719. [doi: 10.1097/NCN.0b013e318224b597] [Medline: 21697705]
17. Hallberg I, Taft C, Ranerup A, Bengtsson U, Hoffmann M, Höfer S, et al. Phases in development of an interactive mobile phone-based system to support self-management of hypertension. Integr Blood Press Control 2014;7:19-28 [FREE Full text] [doi: 10.2147/IBPC.S59030] [Medline: 24910510]
18. Free C, Phillips G, Felix L, Galli L, Patel V, Edwards P. The effectiveness of M-health technologies for improving health and health services: a systematic review protocol. BMC Res Notes 2010;3:250 [FREE Full text] [doi: 10.1186/1756-0500-3-250] [Medline: 20925916]
19. Artinian NT. Telehealth as a tool for enhancing care for patients with cardiovascular disease. J Cardiovasc Nurs 2007;22(1):25-31. [Medline: 17224694]
20. Schwartz KA, Britton B. Use of telehealth to improve chronic disease management. N C Med J 2011;72(3):216-218. [Medline: 21901920]
21. Inglis SC, Clark RA, McAlister FA, Ball J, Lewinter C, Cullington D, et al. Structured telephone support or telemonitoring programmes for patients with chronic heart failure. Cochrane Database Syst Rev 2010(8):CD007228. [doi: 10.1002/14651858.CD007228.pub2] [Medline: 20687083]
22. Dale LP, Whittaker R, Eyles H, Mhurchu CN, Ball K, Smith N, et al. Cardiovascular Disease Self-Management: Pilot Testing of an mHealth Healthy Eating Program. J Pers Med 2014;4(1):88-101 [FREE Full text] [doi: 10.3390/jpm4010088] [Medline: 25562145]
23. Otte-Trojel T, de Bont A, van de Klundert J, Rundall TG. Characteristics of patient portals developed in the context of health information exchanges: early policy effects of incentives in the meaningful use program in the United States. J Med Internet Res 2014;16(11):e258 [FREE Full text] [doi: 10.2196/jmir.3698] [Medline: 25447837]
24. Chow CK, Islam SMS, Farmer A, Bobrow K, Maddision R, Whittaker R, et al. Text2PreventCVD: protocol for a systematic review and individual participant data meta-analysis of text message-based interventions for the prevention of cardiovascular diseases. BMJ Open 2016 Oct 17;6(10):e012723 [FREE Full text] [doi: 10.1136/bmjopen-2016-012723] [Medline: 27798018]
25. Thakkar J, Karthikeyan G, Purohit G, Thakkar S, Sharma J, Verma S, et al. Development of macaronic Hindi-English 'Hinglish' text message content for a coronary heart disease secondary prevention programme. Heart Asia 2016;8(2):32-38 [FREE Full text] [doi: 10.1136/heartasia-2016-010789] [Medline: 27752288]
26. Rawstorn JC, Gant N, Meads A, Warren I, Maddison R. Remotely Delivered Exercise-Based Cardiac Rehabilitation: Design and Content Development of a Novel mHealth Platform. JMIR Mhealth Uhealth 2016;4(2):e57 [FREE Full text] [doi: 10.2196/mhealth.5501] [Medline: 27342791]
27. Park LG, Beatty A, Stafford Z, Whooley MA. Mobile Phone Interventions for the Secondary Prevention of Cardiovascular Disease. Prog Cardiovasc Dis 2016;58(6):639-650. [doi: 10.1016/j.pcad.2016.03.002] [Medline: 27001245]
28. Bashshur RL, Shannon GW, Smith BR, Alverson DC, Antoniotti N, Barsan WG, et al. The empirical foundations of telemedicine interventions for chronic disease management. Telemed J E Health 2014 Sep;20(9):769-800 [FREE Full text] [doi: 10.1089/tmj.2014.9981] [Medline: 24968105]
29. Ammenwerth E, Schnell-Inderst P, Hoerbst A. The impact of electronic patient portals on patient care: a systematic review of controlled trials. J Med Internet Res 2012;14(6):e162 [FREE Full text] [doi: 10.2196/jmir.2238] [Medline: 23183044]
30. Martínez A, Everss E, Rojo-Alvarez JL, Figal DP, García-Alberola A. A systematic review of the literature on home monitoring for patients with heart failure. J Telemed Telecare 2006;12(5):234-241. [doi: 10.1258/135763306777889109] [Medline: 16848935]
31. Seto E. Cost comparison between telemonitoring and usual care of heart failure: a systematic review. Telemed J E Health 2008 Sep;14(7):679-686. [doi: 10.1089/tmj.2007.0114] [Medline: 18817497]
32. Lee M, Lee K, Yoo S, Park C. Impact of initial active engagement in self-monitoring with a telemonitoring device on glycemic control among patients with type 2 diabetes. Sci Rep 2017 Jun 20;7(1):3866 [FREE Full text] [doi: 10.1038/s41598-017-03842-2] [Medline: 28634381]

33. Wen K, Kreps G, Zhu F, Miller S. Consumers' perceptions about and use of the internet for personal health records and health information exchange: analysis of the 2007 Health Information National Trends Survey. J Med Internet Res 2010;12(4):e73 [FREE Full text] [doi: 10.2196/jmir.1668] [Medline: 21169163]
34. Hallberg I, Ranerup A, Kjellgren K. Supporting the self-management of hypertension: Patients' experiences of using a mobile phone-based system. J Hum Hypertens 2016 Feb;30(2):141-146 [FREE Full text] [doi: 10.1038/jhh.2015.37] [Medline: 25903164]
35. Health Information National Trends Survey. About HINTS. Bethesda, MD: National Cancer Institute URL: https://hints.cancer.gov/about-hints/learn-more-about-hints.aspx [WebCite Cache ID 72RV0fbjx]
36. Health Information National Trends Survey. Health Information National Survey 4 (HINTS 4) Cycle 2 Methodology Report. Bethesda, MD: National Cancer Institute; 2013 Mar. URL: https://hints.cancer.gov/docs/Instruments/HINTS_4_Cycle2_Methods_Report.pdf [WebCite Cache ID 72RUypHep]
37. Health Information National Trends Survey. Bethesda, MD: National Cancer Institute. Health Information National Survey 4 (HINTS 4) Cycle 4 Methodology URL: https://hints.cancer.gov/docs/Instruments/HINTS_4_Cycle2_Methods_Report.pdf [WebCite Cache ID 72RUwXNxB]
38. Finney RLJ, Davis T, Beckjord EB, Blake K, Moser RP, Hesse BW. Picking up the pace: changes in method and frame for the health information national trends survey (2011-2014). J Health Commun 2012;17(8):979-989 [FREE Full text] [doi: 10.1080/10810730.2012.700998] [Medline: 23020763]
39. Czaja S, Sharit J, Nair S, Lee C, editors. Older adults internet health information seeking. In: Proceedings of the Human Factors Ergonomics Society Annual Meeting. 2009 Presented at: Human Factors Ergonomics Society Annual Meeting; 2009; San Antonio.
40. Hargittai E. Digital na(t)ives? Variation in internet skills and uses among members of the “net generation”. Sociol Inq 2010;80(1):92-113. [doi: 10.1111/j.1475-682X.2009.00317.x]
41. Martinez VI, Marquard JL, Saver B, Garber L, Preusse P. Consumer Health Informatics Interventions Must Support User Workflows, Be Easy-To-Use, and Improve Cognition: Applying the SEIPS 2.0 Model to Evaluate Patients’ and Clinicians’ Experiences with the CONDUIT-HID Intervention. International Journal of Human-Computer Interaction 2017 Jan 11;33(4):333-343. [doi: 10.1080/10447318.2016.1278340]
42. Isaković M, Sedlar U, Volk M, Bešter J. Usability Pitfalls of Diabetes mHealth Apps for the Elderly. J Diabetes Res 2016;2016:1604609 [FREE Full text] [doi: 10.1155/2016/1604609] [Medline: 27034957]
43. Georgsson M, Staggers N. An evaluation of patients' experienced usability of a diabetes mHealth system using a multi-method approach. J Biomed Inform 2016 Feb;59:115-129 [FREE Full text] [doi: 10.1016/j.jbi.2015.11.008] [Medline: 26639894]
44. Mackert M, Mabry-Flynn A, Champlin S, Donovan EE, Pounders K. Health Literacy and Health Information Technology Adoption: The Potential for a New Digital Divide. J Med Internet Res 2016 Oct 04;18(10):e264 [FREE Full text] [doi: 10.2196/jmir.6349] [Medline: 27702738]
45. Mackert M, Champlin S, Holton A, Muñoz I, Damásio M. eHealth and Health Literacy: A Research Methodology Review. J Comput-Mediat Comm 2014 Apr 12;19(3):516-528. [doi: 10.1111/jcc4.12044] [Medline: 2014]
46. Protheroe J, Whittle R, Bartlam B, Estacio EV, Clark L, Kurth J. Health literacy, associated lifestyle and demographic factors in adult population of an English city: a cross-sectional survey. Health Expect 2017 Feb;20(1):112-119 [FREE Full text] [doi: 10.1111/hex.12440] [Medline: 26774107]
47. Amante DJ, Hogan TP, Pagoto SL, English TM, Lapane KL. Access to care and use of the Internet to search for health information: results from the US National Health Interview Survey. J Med Internet Res 2015;17(4):e106 [FREE Full text] [doi: 10.2196/jmir.4126] [Medline: 25925943]
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;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
49. Widmer R, Allison T, Keane B, Dallas A, Bailey K, Lerman L. Workplace Digital Health Is Associated with Improved Cardiovascular Risk Factors in a Frequency-Dependent Fashion: A Large Prospective Observational Cohort Study. PLoS ONE 2016;11(4):e0152657 [FREE Full text]
50. Kontos EZ, Viswanath K. Cancer-related direct-to-consumer advertising: a critical review. Nat Rev Cancer 2011 Feb;11(2):142-150. [doi: 10.1038/nrc2999] [Medline: 21258398]
51. Fox RC. Cultural competence and the culture of medicine. N Engl J Med 2005 Sep 29;353(13):1316-1319. [doi: 10.1056/NEJMp058066] [Medline: 16192475]
52. Beckjord EB, Finney RLJ, 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]
53. Ramanadhan S, Viswanath K. Health and the information nonseeker: a profile. Health Commun 2006;20(2):131-139 [FREE Full text] [doi: 10.1207/s15327027hc2002_4] [Medline: 16965250]
54. Rutten LJF, Squiers L, Hesse B. Cancer-related information seeking: hints from the 2003 Health Information National Trends Survey (HINTS). J Health Commun 2006;11 Suppl 1:147-156. [doi: 10.1080/10810730600637574] [Medline: 16641080]

Abbreviations

CVD: Cardiovascular disease
eHealth: electronic health
HIE: health information exchange
HINTS: Health Information National Trends Survey
HIT: Health Information Technology
mHealth: mobile health
OR: odds ratio

Edited by G Eysenbach; submitted 16.11.17; peer-reviewed by H Shim, A Beatty, R Widmer; comments to author 15.03.18; revised version received 19.04.18; accepted 10.05.18; published 18.10.18.

Please cite as:
Asan O, Cooper II F, Nagavally S, Walker RJ, Williams JS, Ozieh MN, Egede LE
Preferences for Health Information Technologies Among US Adults: Analysis of the Health Information National Trends Survey
J Med Internet Res 2018;20(10):e277
URL: http://www.jmir.org/2018/10/e277/
doi:10.2196/jmir.9436
PMID:30341048

Original Paper

Modeling and Predicting Outcomes of eHealth Usage by European Physicians: Multidimensional Approach from a Survey of 9196 General Practitioners

Joan Torrent-Sellens[1*], PhD; Ángel Díaz-Chao[2*], PhD; Ivan Soler-Ramos[1*], MSc; Francesc Saigí-Rubió[3*], PhD

[1]Faculty of Economics and Business, Universitat Oberta de Catalunya, Barcelona, Spain
[2]Applied Economics Department, King Juan Carlos University, Madrid, Spain
[3]Faculty of Health Sciences, Universitat Oberta de Catalunya, Barcelona, Spain
[*]all authors contributed equally

Corresponding Author:
Francesc Saigí-Rubió, PhD
Faculty of Health Sciences
Universitat Oberta de Catalunya
Avinguda del Tibidabo 39-43
Barcelona, 08035
Spain
Phone: 34 933263622
Email: fsaigi@uoc.edu

Abstract

Background: The literature has noted the need to use more advanced methods and models to evaluate physicians' outcomes in the shared health care model that electronic health (eHealth) proposes.

Objective: The goal of our study was to design and evaluate a predictive multidimensional model of the outcomes of eHealth usage by European physicians.

Methods: We used 2012-2013 survey data from a sample of 9196 European physicians (general practitioners). We proposed and tested two composite indicators of eHealth usage outcomes (internal practices and practices with patients) through 2-stage structural equation modeling. Logistic regression (odds ratios, ORs) to model the predictors of eHealth usage outcomes indicators were also calculated.

Results: European general practitioners who were female (internal practices OR 1.15, 95% CI 1.10-1.20; practices with patients OR 1.19, 95% CI 1.14-1.24) and younger—aged <35 years (internal practices OR 1.14, 95% CI 1.02-1.26; practices with patients OR 1.32, 95% CI 1.13-1.54) and aged 36-45 years (internal practices OR 1.16, 95% CI 1.06-1.28; practices with patients OR 1.21, 95% CI 1.10-1.33)—had a greater propensity toward favorable eHealth usage outcomes in internal practices and practices with patients. European general practitioners who positively valued information and communication technology (ICT) impact on their personal working processes (internal practices OR 5.30, 95% CI 4.73-5.93; practices with patients OR 4.83, 95% CI 4.32-5.40), teamwork processes (internal practices OR 4.19, 95% CI 3.78-4.65; practices with patients OR 3.38, 95% CI 3.05-3.74), and the doctor-patient relationship (internal practices OR 3.97, 95% CI 3.60-4.37; practices with patients OR 6.02, 95% CI 5.43-6.67) had a high propensity toward favorable effects of eHealth usage on internal practices and practices with patients. More favorable eHealth outcomes were also observed for self-employed European general practitioners (internal practices OR 1.33, 95% CI 1.22-1.45; practices with patients OR 1.10, 95% CI 1.03-1.28). Finally, general practitioners who reported that the number of patients treated in the last 2 years had remained constant (internal practices OR 1.08, 95% CI 1.01-1.17) or increased (practices with patients OR 1.12, 95% CI 1.03-1.22) had a higher propensity toward favorable eHealth usage outcomes.

Conclusions: We provide new evidence of predictors (sociodemographic issues, attitudes toward ICT impacts, and working conditions) that explain favorable eHealth usage outcomes. The results highlight the need to develop more specific policies for eHealth usage to address different realities.

(J Med Internet Res 2018;20(10):e279) doi:10.2196/jmir.9253

KEYWORDS

internet; eHealth; health care; health drivers; health barriers; health attitude; health information; health empowerment; information and communication technologies; structural equation modeling; Europe

Introduction

In recent years, advances in health information technologies (HITs), electronic health records (EHRs), personal health records (PHRs), electronic health (eHealth) and telehealth applications and devices, and health information exchange (HIE) systems have profoundly transformed professionals' health care practice, thus, contributing to the efficiency of their activities by reducing errors, improving early diagnosis of diseases, and offering better management of chronic diseases [1-3]. In the European Union, these HIT and eHealth applications have been incorporated into hospitals and municipalities over the past 10 years, and a higher level of information integration and exchange has been achieved to the benefit of coordination and collaboration between and among health care professionals in all sectors [1,4,5]. In primary care, where 69% of general practitioners use internet-connected computers [1], EHR network usage is well established and widespread [6], whereas HIE system usage is less so [7]. In addition, growing emphasis is being placed on eHealth services as a way of enabling patients to have access to their medical information through PHRs [5,6].

The application of eHealth services to primary care is of particular interest because it is at this level of care where patients come into regular contact with health care systems. Primary care also provides the highest volume of health services [8]. Compared to current practice, however, general practitioners often consider eHealth services as being disruptive rather than supportive [7]. General practitioners are quite skeptical about the potential benefits of eHealth usage to citizens and patients [1] despite the perceived benefits [9]. Moreover, even though the data gathered by different health devices [10] and the Internet of Things [11,12] can serve as a complement to PHRs and help to identify patients' health risks [11], the success of PHRs will be dependent on general practitioners' attitudes and actions. Additional workload coupled with the lack of time, remuneration, information and communication technology (ICT) skills, interoperability, confidentiality, and clear rules about their liability stand out as the main barriers to eHealth usage alongside organizational issues and resistance to change [1,13-17].

While there is considerable evidence in the literature about the predictors of some particular uses of eHealth, attention has recently been drawn to the need to use more advanced methods and models to evaluate the shared health care model that eHealth proposes [18-21]. The construction of a new theoretical framework, a composite indicator of eHealth usage (by patients and health care professionals) that would allow us to understand the integrated and intersectoral workflow and dynamics of eHealth between health care professionals and patients within health care systems would be very useful. Thus, the public policies and strategic actions resulting from the research could be adapted more precisely to specific eHealth uses and the profiles of professionals or health care systems.

Earlier studies have already attempted to model and predict eHealth usage among patients in Europe [22]. The main aim of this work was to model and predict European general practitioners' eHealth usage outcomes. We designed and tested a multidimensional model for this purpose. The results obtained provide new evidence of and have implications for the design of health organizations and public health policies.

Methods

Participants and Procedure

Data for this study were drawn from the Benchmarking Deployment of eHealth among General Practitioners Phase 2 (GPII) research project [23]. The survey was funded by the European Commission Directorate General for Communications Networks, Content and Technology. The GPII panel survey's analysis had two main objectives: (1) measuring the level eHealth availability and usage in primary care and (2) examining what drives or hampers the overall level of eHealth adoption.

The study used survey data from a sample of 9196 European general practitioners, that is to say, physicians working in outpatient establishments in specialties such as general practice, family medicine, internal medicine, or general medicine. The sampling universe comprised 465,718 European physicians with an overall margin of error of ±1.03 in the case of maximum indetermination p=q=50%, for a confidence level of 95.5% (Multimedia Appendix 1). The sample had two essential characteristics. First, the survey selected a large sample of countries (31 in total): 27 European Union countries (Austria, Belgium, Bulgaria, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, The Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, and United Kingdom) plus Croatia, Iceland, Norway, and Turkey. Second, it selected an unequal-sized sample with country-specific margins of error varying between +4.15% and +13.84%. The margins of error for most of the countries in the sample (n=20) were around ±5.0%, whereas for 6 countries they were between ±6.0% and ±7.0%, and for only 5 countries (the smaller ones), they were above ±10.0%.

The questionnaire used in the survey contained 38 questions grouped into 3 dimensions (Multimedia Appendix 2). Part A covered general practitioners' sociodemographic circumstances, organizational settings, practice location, and description of tasks and workload (10 questions). Part B covered the deployment and usage of ICT systems and functionalities (23 questions) and represented the core of the survey. After a set of general questions (basic infrastructure, interconnection with other system players, and security items), the following 4 pillars of eHealth usage measurement were addressed in this dimension: (1) EHRs, referring to the systems that are used by health care professionals to enter, store, view, and manage patient health and administrative information and data; (2) HIEs, referring to the process of electronically transferring, sharing, or enabling access to patient health information and data; (3) telehealth, covering the use of broadband-based technological platforms for the purpose of providing health services, medical training, and health education at a distance; and (4) PHRs, referring to the electronic systems allowing patients to have secure access to and manage their health information. Finally, Part C focuses on attitudes toward, perceived barriers to, and impacts of eHealth usage (5 questions).

The survey was answered by European general practitioners in mixed Web-based, phone (Web-CATI), and face-to-face interviews lasting for half an hour each and in a native language of each country. A study presentation paragraph was written to inform potential respondents about the confidentiality of any data provided and the academic aim of the research. European general practitioners voluntarily answered the questionnaire and did not receive any payment in cash or kind. While the questionnaire was being implemented, an expert was on hand at all times (via email) to resolve any queries that the respondents had. The respondent general practitioners were selected by means of probability sampling applied to each country universe. The net response rate was 35.5%. The fieldwork period ran from October 25, 2012 to March 6, 2013. The GPII research project followed the Checklist for Reporting Results of Internet E-Surveys criteria [24]. For a more detailed explanation, see the GPII research report [23].

Data Analysis and Models

From an empirical perspective, explanatory factors determining eHealth usage outcomes raise two particular difficulties. First, the approach to the concept requires a multidimensional basis that is not usually captured in a single variable. In fact, the most common approaches found in the literature perform partial analyses of its various dimensions. This type of analysis has the disadvantage of not taking a full snapshot of the explanatory factors, which gives rise to the second difficulty: statistical modeling. In other words, eHealth usage outcomes can be interpreted as a latent, nonobservable concept, which, therefore, calls for statistical techniques that allow variables of this type, which are not directly measurable, to be used [22].

In the empirical literature, structural equation modeling (SEM) with latent variables has been used to overcome this problem. A general SEM is a formal mathematical model. It is a set of linear equations that encompasses various types of models such as regression analysis models, simultaneous equation systems, factor analysis, and path analysis. The main advantage of this method of analysis is the incorporation of different types of variables into the SEM. Directly observable and measurable variables, and theoretical or latent variables representing concepts that are not directly observed can, therefore, be incorporated. When the variable to be explained (dependent) is latent, it must be continuous, whereas dependent observed variables can be continuous, censored, binary, ordered, categorical (ordinals), or combinations of any of these variable types [25].

This method of analysis allows us to define eHealth usage outcomes as a latent variable, thus enabling us to calculate the specific explanatory effect of the variables that it comprises. Hence, besides building an overall explanatory model of the determinants of eHealth usage, it is also possible to identify which of its explanatory dimensions are more important. In addition, SEM enables the relationships between the different observable variables included in the model (indirect effects) to be estimated. In this context, and in order to capture the factors that explain eHealth usage outcomes in a large sample of European general practitioners, we proposed and tested a two-stage SEM with latent variables and measurement errors for 2012-2013.

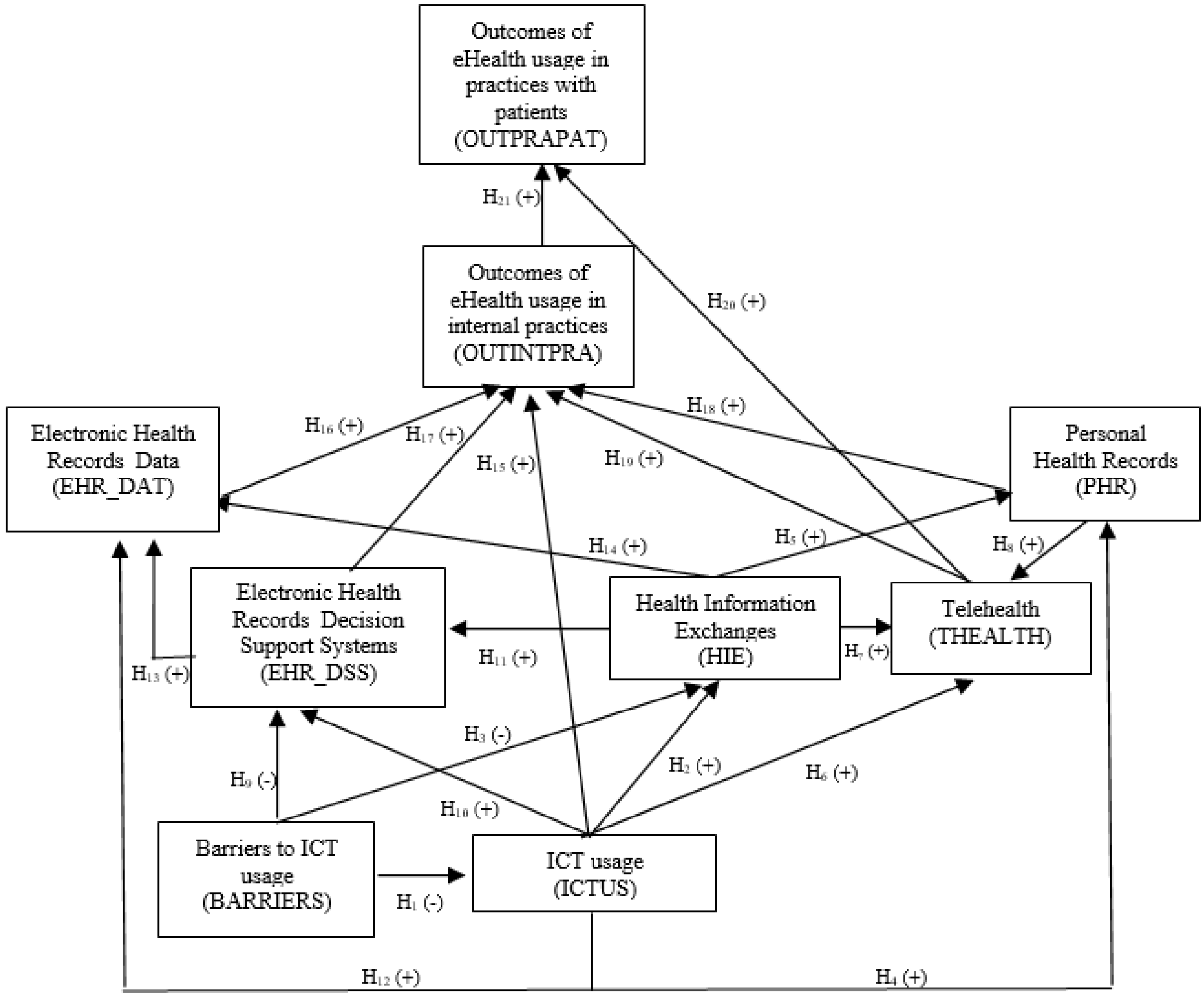
We applied the 2-stage empirical methodology as follows: in the first stage, we tested the relationships among 101 indicators and the 9 dimensions describing eHealth usage and eHealth usage outcomes by means of SEM and additive indicators (in those dimensions with primary data of a dichotomous nature), and in the second stage, we tested the relationships among the indicators constructed for those 9 dimensions (based on the coefficients and aggregations from the first stage). This methodology involved the design and statistical testing of 5 empirical SEM models (4 models for the first stage and 1 model for the second stage) and also 5 additive indicators in the first stage.

The 9 model dimensions and variables are as follows:

- *Dimension 1: ICT usage (ICTUS)*, captured by a set of 5 variables measuring usage frequency (Multimedia Appendix 3);
- *Dimension 2: barriers to eHealth usage (BARRIERS)*, captured by a set of 16 variables measuring the factors that general practitioners regarded as barriers when evaluating eHealth usage (Multimedia Appendix 4);
- *Dimension 3: PHR usage (PHR)*, captured by a set of 6 variables measuring their usage or nonusage (Multimedia Appendix 5);
- *Dimension 4: telehealth (THEALTH)*, captured by a set of 4 variables measuring its usage or nonusage (Multimedia Appendix 6);
- *Dimension 5: HIE*, captured by a set of 15 variables measuring their usage or nonusage (Multimedia Appendix 7);
- *Dimension 6: Electronic Health Records_Decision Support Systems (EHR_DSS)*, captured by a set of 6 variables measuring their usage or nonusage (Multimedia Appendix 8);
- *Dimension 7*: *Electronic Health Records_Data (EHR_DAT)* captured by a set of 19 variables measuring their usage or nonusage (Multimedia Appendix 9);
- *Dimension 8: eHealth usage outcomes in internal practices (OUTINTPRA)*, captured by a set of 14 variables measuring the outcomes that general practitioners considered when evaluating eHealth usage in their internal practices (IP; Multimedia Appendix 10);
- *Dimension 9: eHealth usage outcomes in practices with patients (OUTPRAPAT)*, captured by a set of 16 variables measuring the outcomes that general practitioners considered when evaluating eHealth usage in their practices with patients (PP; Multimedia Appendix 11).

Figure 1 shows the multidimensional model of eHealth usage dimensions and outcomes that we have tested.

Figure 1. Model of eHealth usage and eHealth usage outcomes. H: hypothesis; ICT: information and communication technology; (+): positive prediction; (−) negative prediction.



Additionally, we performed logistic regression to model the predictors of 2 eHealth usage outcomes indicators using independent variables corresponding to sociodemographic circumstances, attitudes, and working conditions. For each independent variable, we calculated odds ratios (ORs) and their 95% CI. We used IBM SPSS Amos v24 (IBM Corp) for all calculations.

Results

eHealth Usage Outcomes

Multimedia Appendix 12 shows the results (standardized coefficients and measurement errors) of the first stage of estimating the explanatory factors of some dimensions of European general practitioners' eHealth usage and of 2 dimensions of general practitioners' eHealth usage outcomes for 2012-2013. In this first stage, we estimated the relationships among 51 indicators and 4 dimensions describing eHealth usage (ICTUS and BARRIERS) and eHealth usage outcomes (OUTINTPRA and OUTPRAPAT) using an SEM with measurement errors. First, it should be noted that all the variables specified in the model were statistically significant (99% confidence level). Second, the goodness-of-fit measurements for the 4 proposed models were highly satisfactory. Thus, the normed fit index (NFI), relative fit index (RFI), incremental fit index (IFI), Tucker-Lewis index (TLI), and comparative fit index (CFI) had very high values, approaching the optimal value of 1. The root mean square error of approximation (RMSEA) values were <0.08, thus, corroborating the validity of the estimated models.

In the ICTUS dimension, the standardized coefficient variability was 0.58. The variables with the highest explanatory power in this dimension were related to describing the medical organizations in contact with general practitioners (0.640) as well as the existence of problems of compatibility in electronically exchanging patient data (0.481). In contrast, fewer explanatory variables were related to computer usage in general practice (0.064). In the BARRIERS dimension, the standardized coefficient variability was much lower (0.15) between the explanatory variables related to the lack of time or additional workload (0.681), lack of sufficient training (0.673) or lack of sufficient ICT skills (0.663), and increased patient expectations (0.528).

Regarding the 2 dimensions of general practitioners' eHealth usage outcomes, the standardized coefficients obtained showed a different variability. In the case of eHealth usage outcomes related to IP, the standardized coefficient variability was 0.37. The variables with the highest explanatory power were related to the fact that eHealth was easy to use (0.801), that general practitioners found it easy to get it to do what they wanted (0.769), and that it was flexible to use or interact with (0.744). The variables with the least explanatory power were related to people who influence general practitioners' general behavior (0.442) or clinical behavior (0.435) in respect of eHealth usage. In the case of eHealth usage outcomes in PP, the standardized coefficient variability was much lower (0.15). The variables with the highest explanatory power were related to improvement in the efficiency of the whole health care system (0.797), in the quality of treatment (0.784), and in the quality of diagnosis decisions (0.783). In contrast, the variables with the least explanatory power were related to the general practitioners' perceptions of eHealth usage increasing patient access to health care (0.647) or reducing pharmaceutical expenditure (0.649).

For the remaining 5 dimensions relating to eHealth usage, namely PHR, THEALTH, HIEs, EHR_DSS, and EHR_DAT, and as a result of the dichotomous nature of the base data, we constructed 5 additive indicators. For each of the 5 dimensions, we created a joint indicator that adds together their internal uses (indicators). Thus, the PHR indicator takes a value between 0 and 6, the THEALTH indicator between 0 and 4, the HIE indicator between 0 and 15, the EHR_DSS indicator between 0 and 6, and the EHR_DAT indicator between 0 and 19 (Multimedia Appendix 13). Some 37.58% (3456/9196) and 41.90% (3853/9196) of the general practitioners performed at least 1 of the 6 PHR practices and of the 4 THEALTH practices. However, the percentages in the HIE and EHR (DSS and DAT) dimensions were much higher. Some 87.99% (8092/9196), 70.20% (6456/9196), and 81.89% (7531/9196) of European general practitioners made at least 1 use of eHealth in the HIE and EHR (DSS and DAT) dimensions. After applying the coefficients—obtained from the first estimation stage—and the additive indicators, we constructed 7 composite indicators of eHealth usage and 2 of eHealth usage outcomes and determined their mean values (Table 1).

eHealth Usage Outcomes in Internal Practices and in Practices With Patients

Table 2 shows the results (direct standardized coefficients and standard errors) of the second stage of modeling European general practitioners' eHealth usage and eHealth usage outcomes for 2012-2013. In this second stage, we tested the relationships among the indicators constructed for the 7 dimensions describing eHealth usage (based on the coefficients and additive indicators from the first stage) and the 2 latent constructs of eHealth usage outcomes (based on the coefficients from the first stage) by using a 21-hypothesis SEM with standard errors (Figure 1). First, it should be noted that all the variables specified in the model were statistically significant (at least 95% confidence level). Second, the goodness-of-fit measurements for the proposed model were satisfactory. Thus, the NFI (0.966), RFI (0.914), IFI (0.967), TLI (0.915), and CFI (0.967) indices had high values, approaching the optimal value of 1. The RMSEA value was <0.08 (0.072), thus, corroborating the validity of the estimated model.

The direct standardized coefficients obtained validated the 21 formulated hypotheses and the signs of their relationships. For example, as we had hypothesized, the barriers to ICTUS determined a negative effect on ICT, HIE, and EHR_DSS usage. Regarding the relationships between different dimensions of eHealth usage, it is notable that ICTUS explained HIE (0.571) with considerable intensity that HIE had considerable explanatory power over PHR (0.398) and EHR_DSS (0.349) and that an important link was found between EHR_DSS and EHR_DAT (0.365). Concerning the indicators relating to eHealth usage outcomes, the indicator linked to IP was explained by ICT (0.132), EHR_DAT (0.215), EHR_DSS (0.050), PHR (0.112) and THEALTH (0.101) usage. In contrast, the outcomes indicator linked to PP was explained by THEALTH usage (0.027) and, primarily, by the IP outcomes indicator (0.607).

Table 1. Descriptive statistics of eHealth usage and eHealth usage outcome dimensions, 2012-2013.

#	Dimension	Mean (SD)	Minimum	Maximum	Skewness	Kurtosis
1	Information and communication technology usage	1.54 (1.002)	0.00	4.54	0.085	0.738
2	Barriers to eHealth usage	9.99 (3.014)	0.00	15.34	−0.855	1.177
3	Personal Health Records	0.82 (1.320)	0.00	6.00	1.872	3.360
4	Telehealth	0.62 (0.865)	0.00	4.00	1.450	1.747
5	Health Information Exchanges	5.26 (3.962)	0.00	15.00	0.551	−0.569
6	Electronic Health Records_Decision Support Systems	2.38 (2.127)	0.00	6.00	0.371	−1.233
7	Electronic Health Records_Data	13.80 (6.015)	0.00	19.00	−1.428	0.628
8	eHealth usage outcomes in internal practices	8.86 (2.719)	0.00	13.12	−0.785	0.681
9	eHealth usage outcomes in practices with patients	11.50 (3.712)	0.00	16.88	−0.785	0.741

Table 2. Explanatory factors of European general practitioners' eHealth usage and eHealth usage outcomes, 2012-2013. Goodness-of-fit indices: normed fit index: 0.966; relative fit index: 0.914; incremental fit index: 0.967; Tucker-Lewis index: 0.915; comparative fit index: 0.967; root mean square error of approximation: 0.072.

Hypothesis (H)	Explained dimension	Explanatory dimension	Standardized coefficient[a]	*P* value of standardized coefficients	SE	*P* value of SEs
H1	ICTUS[b]	BARRIERS[c]	−0.118	<.001	0.003	<.001
H2	HIE[d]	ICTUS	0.571	<.001	0.034	<.001
H3	HIE	BARRIERS	−0.035	<.001	0.011	<.001
H4	PHR[e]	ICTUS	0.062	<.001	0.015	<.001
H5	PHR	HIEs	0.398	<.001	0.004	<.001
H6	THEALTH[f]	ICTUS	0.031	<.001	0.010	.003
H7	THEALTH	HIEs	0.248	<.001	0.003	<.001
H8	THEALTH	PHR	0.081	<.001	0.007	<.001
H9	EHR_DSS[g]	ICTUS	0.191	<.001	0.023	<.001
H10	EHR_DSS	BARRIERS	−0.012	<.001	0.006	.003
H11	EHR_DSS	HIEs	0.349	<.001	0.006	<.001
H12	EHR_DAT[h]	ICTUS	0.190	<.001	0.062	<.001
H13	EHR_DAT	EHR_DSS	0.365	<.001	0.027	<.001
H14	EHR_DAT	HIEs	0.174	<.001	0.016	<.001
H15	OUTINTPRA[i]	ICTUS	0.132	<.001	0.030	<.001
H16	OUTINTPRA	EHR_DAT	0.215	<.001	0.005	<.001
H17	OUTINTPRA	EHR_DSS	0.050	<.001	0.015	<.001
H18	OUTINTPRA	PHR	0.112	<.001	0.021	<.001
H19	OUTINTPRA	THEALTH	0.101	<.001	0.031	<.001
H20	OUTPRAPAT[j]	THEALTH	0.027	.002	0.036	.003
H21	OUTPRAPAT	OUTINTPRA	0.607	<.001	0.011	<.001

[a] Regression analysis: structural equation modeling (SEM); estimated coefficients: direct effects.

[b] ICTUS: information and communication technology usage.

[c] BARRIERS: barriers to information and communication technology usage.

[d] HIE: Health Information Exchange.

[e] PHR: Personal Health Records.

[f] THEALTH: telehealth.

[g] EHR_DSS: Electronic Health Records_Decision Support Systems.

[h] EHR_DAT: Electronic Health Records_Data.

[i] OUTINTPRA: eHealth usage outcomes in internal practices.

[j] OUTPRAPAT: eHealth usage outcomes in practices with patients.

Predictors of eHealth Usage Outcomes in Internal Practices

To capture the main predictors of European general practitioners' eHealth usage outcomes, we performed logistic regression using independent variables for general practitioners' sociodemographic circumstances, attitudes toward ICT impact, and working conditions. The first step in this analysis was to recode the two composite indicators of eHealth usage outcomes. The mean value the composite indicator of OUTINTPRA was 8.86 (SD 2.72; minimum to maximum range 0.0-13.12). The mean value of the composite indicator of OUTPRAPAT was 11.5 (SD 3.71; minimum to maximum range 0.0-16.88). We, therefore, constructed a dichotomous eHealth usage outcomes indicator based on the mean of the composite indicators obtained. The two dichotomous eHealth usage outcomes indicators took the value 1 when the composite indicators of eHealth usage outcomes were equal to or greater than the mean and the value 0 when less than the mean. The mean value of the dichotomous composite indicator of OUTINTPRA was 0.547 (SD 0.498; minimum to maximum range 0-1). The mean value of the dichotomous indicator of OUTPRAPAT was 0.558

(SD 0.497; minimum to maximum range 0-1). For 54.75% (5035/9196) of European general practitioners, the eHealth usage outcomes in IP were more favorable (greater than the mean). For 55.82% (5133/9196) of European general practitioners, the eHealth usage outcomes in PP were more favorable (greater than the mean).

Table 3 shows the results of the logistic regression (OR) between the dichotomous composite indicator of OUTINTPRA and the independent variables. Regarding sociodemographic circumstances, European general practitioners who were female (OR 1.15, 95% CI 1.10-1.20) and younger (aged <35 years: OR 1.14, 95% CI 1.02-1.26 and aged 36-45 years: OR 1.16, 95% CI 1.06-1.28) had a higher propensity toward favorable OUTINTPRA than male general practitioners (OR 0.89, 95% CI 0.85-0.92) or those in older age groups (aged 46-55 years: OR 0.95, 95% CI 0.87-1.03 and aged >56 years: OR 0.90, 95% CI 0.82-0.98). We found no significant differences between European general practitioners' workplace location and more favorable OUTINTPRA: large cities (>100,000 inhabitants): OR 1.02, 95% CI 0.93-1.11, small-midsized cities (20,000-100,000 inhabitants): (OR 0.98, 95% CI 0.89-1.07), and rural towns (<20,000 inhabitants): OR 1.00, 95% CI 0.97-1.03).

Concerning attitudes toward perceived ICT impact on IP, the results obtained showed considerable predictive power. European general practitioners who positively valued ICT impact on their personal working processes (OR 5.30, 95% CI 4.73-5.93), teamwork processes (OR 4.19, 95% CI 3.78-4.65), and the doctor-patient relationship (OR 3.97, 95% CI 3.60-4.37) had a high propensity toward favorable effects of OUTINTPRA. In contrast, European general practitioners who negatively valued the effect of ICT impact or considered that it had no effect on their personal working processes (no effect: OR 0.26, 95% CI 0.23-0.29; negatively: OR 0.12, 95% CI 0.09-0.15), teamwork processes (no effect: OR 0.31, 95% CI 0.28-0.35; negatively: OR 0.17, 95% CI 0.14-0.22), or the doctor-patient relationship (no effect: OR 0.51, 95% CI 0.47-0.56; negatively: OR 0.27, 95% CI 0.24-0.31) had a lower propensity toward favorable effects of eHealth usage. A clear link was also observed between a greater personal use of Web 2.0 (social media, blogs, etc) and more favorable OUTINTPRA. European general practitioners who had used 2.0 applications in their personal lives often (OR 1.77, 95% CI 1.60-1.97) or sometimes (OR 1.16, 95% CI 1.06-1.28) also showed a higher propensity toward favorable OUTINTPRA than those who rarely (OR 0.99, 95% CI 0.89-1.09) or never (OR 0.59, 95% CI 0.55-0.65) used 2.0 applications.

Finally, the study also found predictive power between some working conditions and favorable OUTINTPRA. Self-employed European general practitioners (OR 1.33, 95% CI 1.22-1.45) also had a greater predisposition toward favorable eHealth usage outcomes in IP than salaried European general practitioners (OR 0.82, 95% CI 0.75-0.89). For its part, the dynamic of the number of patients treated also had predictive power. In the case of favorable OUTINTPRA, the results obtained had an inverted U shape. European general practitioners who reported that the number of patients treated had remained constant in the last 2 years (OR 1.08, 95% CI 1.01-1.17) had a greater predisposition toward favorable eHealth usage outcomes in their IP. In contrast, general practitioners who reported that the number of patients treated had increased (1.03, 95% CI 0.95-1.12) or had fallen in recent years (OR 0.81, 95% CI 0.72-0.91) had less predictive power.

Predictors of eHealth Usage Outcomes in Practices With Patients

Table 4 shows the results of the logistic regression (OR) between the dichotomous composite indicator of OUTPRAPAT and the independent variables. Regarding sociodemographic circumstances, European general practitioners who were female (OR 1.19, 95% CI 1.14-1.24) and younger (aged <35 years: OR 1.32, 95% CI 1.13-1.54 and aged 36-45 years: OR 1.21, 95% CI 1.10-1.33) had a higher propensity toward favorable OUTPRAPAT than male general practitioners (OR 0.86, 95% CI 0.83-0.89) or those in older age groups (aged 46-55 years: OR 0.84, 95% CI 0.77-0.92 and aged >56 years: OR 0.93, 95% CI 0.85-1.01). We found no significant differences between European general practitioners' workplace location and more favorable OUTPRAPAT: large cities (>100,000 inhabitants): OR 0.97, 95% CI 0.89-1.05; small or midsized cities (20,000-100,000 inhabitants): OR 1.00, 95% CI 0.91-1.09; and rural towns (<20,000 inhabitants): OR 1.04, 95% CI 0.95-1.13.

Concerning attitudes toward perceived ICT impact on IP, the results obtained showed considerable predictive power. European general practitioners who positively valued ICT impact on their personal working processes (OR 4.83, 95% CI 4.32-5.40), teamwork processes (OR 3.38, 95% CI 3.05-3.74), and the doctor-patient relationship (OR 6.02, 95% CI 5.43-6.67) had a high propensity toward favorable effects of eHealth usage on PP. In contrast, European general practitioners who negatively valued the effect of ICT impact or considered that it had no effect on their personal working processes (no effect: OR 0.26, 95% CI 0.23-0.30; negatively: OR 0.17, 95% CI 0.13-0.21), teamwork processes (no effect: OR 0.38, 95% CI 0.34-0.42); negatively: OR 0.20, 95% CI 0.16-0.26), and the doctor-patient relationship (no effect: OR 0.38, 95% CI 0.35-0.42; negatively: OR 0.26, 95% CI 0.22-0.29) had a lower propensity toward favorable effects of eHealth usage. A clear link was also observed between a greater personal use of Web 2.0 (social media, blogs, etc) and more favorable OUTPRAPAT. European general practitioners who had used 2.0 applications in their personal lives often (OR 1.94, 95% CI 1.74-2.15) or sometimes (OR 1.19, 95% CI 1.08-1.31) also showed a higher propensity toward favorable OUTPRAPAT than those who rarely (OR 1.09, 95% CI 0.98-1.20) or never (OR 0.51, 95% CI 0.47-0.56) used 2.0 applications.

Table 3. Logistic regression (odds ratio, OR) models of the dichotomous composite indicator of eHealth usage outcomes in internal practices by sociodemographic circumstances, attitudes toward information and communication technology (ICT) impact, and working conditions, 2012-2013.

Sociodemographic factors	OR	95% CI
Gender		
Male	0.89	0.85-0.92
Female	1.15	1.10-1.20
Age range (years)		
<35	1.14	1.02-1.26
36-45	1.16	1.06-1.28
46-55	0.95	0.87-1.03
>56	0.90	0.82-0.98
Workplace location		
Large city (more than 100,000 inhabitants)	1.02	0.93-1.11
Small or midsized city (between 20,000 and 100,000 inhabitants)	0.98	0.89-1.07
Rural town (fewer than 20,000 inhabitants)	1.00	0.97-1.03
ICT impact on personal working processes		
Positive	5.30	4.73-5.93
No change	0.26	0.23-0.29
Negative	0.12	0.09-0.15
ICT impact on teamwork processes		
Positive	4.19	3.78-4.65
No change	0.31	0.28-0.35
Negative	0.17	0.14-0.22
ICT impact on the doctor-patient relationship		
Positive	3.97	3.60-4.37
No change	0.51	0.47-0.56
Negative	0.27	0.24-0.31
Web 2.0 (social media, blogs, etc) usage in personal life		
Often	1.77	1.60-1.97
Sometimes	1.16	1.06-1.28
Rarely	0.99	0.89-1.09
Never	0.59	0.55-0.65
Occupational status		
Salaried	0.82	0.75-0.89
Self-employed	1.33	1.22-1.45
Number of patients treated in the last 2 years		
Increased	1.03	0.95-1.12
Remained constant	1.08	1.01-1.17
Decreased	0.81	0.72-0.91

Table 4. Logistic regression (odds ratio, OR) models of the dichotomous composite indicator of eHealth usage outcomes in practices with patients by sociodemographic circumstances, attitudes toward information and communication technology (ICT), and working conditions, 2012-2013.

Sociodemographic factors	OR	95% CI
Gender		
Male	0.86	0.83-0.89
Female	1.19	1.14-1.24
Age range (years)		
<35	1.32	1.13-1.54
36-45	1.21	1.10-1.33
46-55	0.84	0.77-0.92
>56	0.93	0.85-1.01
Workplace location		
Large city (>100,000 inhabitants)	0.97	0.89-1.05
Small or midsized city (20,000-100,000 inhabitants)	1.00	0.91-1.09
Rural town (<20,000 inhabitants)	1.04	0.95-1.13
ICT impact on personal working processes		
Positive	4.83	4.32-5.40
No change	0.26	0.23-0.30
Negative	0.17	0.13-0.21
ICT impact on teamwork processes		
Positive	3.38	3.05-3.74
No change	0.38	0.34-0.42
Negative	0.20	0.16-0.26
ICT impact on the doctor-patient relationship		
Positive	6.02	5.43-6.67
No change	0.38	0.35-0.42
Negative	0.26	0.22-0.29
Web 2.0 (social media, blogs, etc) usage in personal life		
Often	1.94	1.74-2.15
Sometimes	1.19	1.08-1.31
Rarely	1.09	0.98-1.20
Never	0.51	0.47-0.56
Occupational status		
Salaried	1.02	0.93-1.11
Self-employed	1.10	1.03-1.28
Number of patients treated in the last 2 years		
Increased	1.12	1.03-1.22
Remained constant	0.93	0.85-1.01
Decreased	0.91	0.81-1.03

Finally, the study also found predictive power between some working conditions and favorable OUTPRAPAT. Self-employed European general practitioners (OR 1.10, 95% CI 1.03-1.28) also had a greater predisposition toward favorable OUTPRAPAT than salaried European general practitioners (OR 1.02, 95% CI 0.93-1.11). For its part, the dynamic of the number of patients treated also had predictive power. In the case of favorable OUTPRAPAT, the results obtained had a growing evolution. European general practitioners who reported that the number of patients treated had remained constant in the last 2 years (OR 1.12, 95% CI 1.03-1.22) had a greater predisposition toward favorable OUTPRAPAT. In contrast, general practitioners who reported that the number of patients treated had remained constant (OR 0.93, 95% CI 0.85-1.01) or had fallen (OR 0.91, 95% CI 0.81-1.03) did not have predictive power.

Discussion

Principal Findings

The goal of our study was to design and evaluate a predictive multidimensional model of general practitioners' eHealth usage outcomes in IP and in PP, comprising 9 dimensions and 101 indicators. To that end, we used a broad population sample of 9196 European general practitioners. The results obtained are very useful for two reasons. First, obtaining new evidence centered solely on general practitioners allowed us to focus the analysis better, particularly with regard to the eHealth usage dimensions (ICTUS, BARRIERS, PHRs, HIEs, THEALTH, and EHRs) that determine favorable eHealth usage outcomes. Second, the predictors we obtained (sociodemographic circumstances, attitudes toward ICT impact, and working conditions) provided evidence that complements studies that have taken partial approaches.

eHealth Usage Outcomes Indicators

In our study, we constructed 2 composite indicators using a 2-stage SEM methodology, and the results obtained are consistent with this evidence: they showed that in 2012-2013: (1) for 54.75% (5035/9196) of European general practitioners, the eHealth usage outcomes in IP were more favorable (greater than the mean) and (2) for 55.82% (5133/9196) of European general practitioners, the eHealth usage outcomes in PP were more favorable (greater than the mean).

The outcomes indicator linked to IP were explained by ICTUS, EHR_DAT, EHR_DSS, PHRs, and THEALTH. In contrast, the outcomes indicator linked to PP were explained by THEALTH usage and, primarily, by the IP outcomes indicator. General practitioners' eHealth usage outcomes had a 2-fold interrelated dynamic. Firstly, the set of eHealth uses had explanatory power over IP outcomes. And secondly, the IP outcomes, together with THEALTH usage, determined eHealth usage outcomes in PP. In the explanation of general practitioners' eHealth usage, it, therefore, seems that a certain "experience effect" occurs. eHealth usage takes place initially in IP and is then transferred to PP. In addition, the results of this study revealed the key role that telemedicine would play in the sense that, through training programs, upskilling, and learning, it would enable eHealth usage to be transferred from IP to PP [26].

Our study confirms that the perception of eHealth being easy to use, of general practitioners finding it easy to get it to do want they want, and of it being flexible to use or interact with are explanatory factors that have a bigger effect on eHealth usage in IP. The statistical significance of these determining factors refers back to the importance of perceived usefulness and ease of use when the use of a technology needs to be explained. General practitioners' surgeries are characterized by the high number of patients cared for. Patients have various health problems, some of which are clinically complex. As a result, general practitioners have to handle several aspects at the same time, which may lead to doubts in their daily clinical practice [27]. That is why they need systems that are easy and flexible to use as well as being useful [28]. Hence the addition of "flexibility" to these two terms. The need to optimize time within a context of cutbacks in health care provision and spending may explain the significance of these determinants relating to general practitioners' internal activities. Thus, if technology is adapted to the local context, does not entail any added complexity [29], and facilitates real-time access to data that is both reliable and fast, it will have an influence on its acceptance and adoption [28].

Regarding eHealth usage outcomes in PP, the variables with the highest explanatory power were related to improvement in the efficiency of the whole health care system, in the quality of treatment (0.784) and in the quality of diagnosis decisions. In this respect, eHealth is perceived as a technology that serves to reduce costs and increase the quality of health care provision [30-33]. General practitioners' positive attitudes toward ICTUS explained and increased HIE usage. HIE had important explanatory power over PHR and EHR_DSS, and an important link between EHR_DSS and EHR_DAT was also found. Experiences that draw on the advantages of HIE and telemedicine to improve communication between general practitioners and their colleagues or specialists have been shown to be beneficial in terms of efficiency, cost-effectiveness, and improved medical care, with a high degree of satisfaction [32,34,35]. General practitioners' adoption of EHR_DAT is crucial to compiling information across the entire health care system because they are the first point of contact in the provision of integrated health care supported by the potential of eHealth [36,37]. In addition, PHRs enable them to deal with all the patient's self-reported information within the limited amount of time allocated to a clinical visit [38].

The variables with less explanatory power were related to general practitioners' perceptions of whether or not eHealth usage increases patient access to health care or reduces pharmaceutical expenditure. Again, the variables with less explanatory power were related to the belief that electronic systems would disrupt health care provision[30,39]. It is important to ensure that the potential benefits of new technologies are clear to see within the organization through mechanisms of continuing evaluation and feedback [40].

Predictors of eHealth Usage Outcomes (Internal Practices and Practices With Patients)

General practitioner's eHealth usage in IP and in PP has been shown to have significant relationships with the individual characteristics of general practitioners, such as gender and age. It has been described in the literature that demographic factors such as age, education, gender, nationality, and clinical experience can have an influence on health care professionals' predisposition toward eHealth systems [8,41,42].

Concerning attitudes toward the perceived ICT impact on internal and external practices, our analysis revealed that European general practitioners who positively valued ICT impact on their personal working processes, teamwork processes, and the doctor-patient relationship (IP OR 3.97, 95% CI 3.60-4.37) had a high propensity toward favorable OUTINTPRA. A clear link was also observed between a greater personal use of Web 2.0 (social media, blogs, etc) and more favorable OUTINTPRA. Various studies have described organizational and individual barriers to the implementation of eHealth services. Recognizing and understanding what the barriers and facilitators are is ideal for devising strategies and interventions to improve the effective eHealth usage and to address the barriers to implementation [43].

Regarding occupational status, our study has also revealed that self-employed European general practitioners had a greater predisposition toward favorable OUTINTPRA and OUTPRAPAT than salaried European general practitioners. Specifically, self-employed general practitioners attached importance to the potential of eHealth to reduce costs [30,32], despite the fact that practices with a single general practitioner might come up against higher barriers (eg, the costs of buying and maintaining technology) and face difficulties in terms of securing access to other essential resources for the implementation and continued use of eHealth services (eg, information technology support and training time) [44,45].

Last, from the perspective of demand pressures on health care practices, European general practitioners who reported that the number of patients treated had remained constant in the last 2 years had a greater predisposition toward favorable eHealth usage outcomes in their IP. In contrast, general practitioners who reported that the number of patients treated had increased or had fallen had less predictive power. This would, therefore, confirm the available evidence, which associates the implementation of digital technologies in the health care field with specific organizational circumstances, in particular with workflow pressures that are not too heavy [39,46,47].

In contrast—and this is the only result that clearly distinguishes between eHealth usage for IP and that for PP—demand pressures would have predictive power over positive eHealth outcomes. European general practitioners who reported that the number of patients treated had remained constant in the last 2 years had a greater predisposition toward favorable OUTPRAPAT [30,33,47]. While positive eHealth usage in IP was associated with demand that remained constant, in PP it would be associated with growing demand. Once again, it seems that the training effect prevails in the sense that eHealth usage is tested initially in IP, without the pressure of greater attention, and is then transferred to health care PP that are more pressurized.

Limitations

Our study has several limitations. First, there was a time lag between the years the data were obtained (2012-2013) and the year we wrote the paper. However, we felt that the availability of a population database of 9196 European general practitioners deserved an analysis despite the time lag. In future research, and as they become available, we will use newer data and introduce dynamic comparisons. Second, the study provides information only from the perspective of physicians. In the future, we intend to address the issue of eHealth usage by health professionals and health users. By doing so, we will be able to improve our multidimensional approach and obtain mixing results and conclusions for all actors involved in eHealth usage and eHealth usage outcomes. Third, the empirical methodology could also be improved by looking at the intensity of eHealth usage (not simply usage or mean usage) and at a higher number of predictors.

Conclusions

The results obtained highlight the need for more in-depth research to be conducted into the link between eHealth usage, eHealth outcomes and predictors, and the different health care systems in Europe. By doing so, it will be possible to increase the resolution of our results and to establish whether the intensity of eHealth usage and eHealth outcomes varies depending on the health care systems or the extent to which health care systems determine the prediction of eHealth usage or eHealth outcomes. Similarly, strategic and public policy actions resulting from the research could be adapted more precisely to each health care system. Finally, the study results could be supplemented by the construction of a composite indicator of eHealth usage by health care professionals and health care users. The design, validation, and prediction of composite indicators of eHealth usage and eHealth outcomes that take into consideration the perspectives of both users (ie, patients) and professionals in the different European health care systems would provide us with a very comprehensive view of the issue and would allow us to round off our multidimensional approach. We shall focus our efforts on all of these approaches in the near future.

Again, appropriate innovations are needed to promote eHealth usage. European, national and regional authorities should take the results of these studies into account to develop suitable policies for greater integration of HIT among European health care professionals. This setting poses significant challenges for the formulation of public policies and strategies by states where decisions about eHealth should not be overlooked.

Acknowledgments

This work used data from the Benchmarking Deployment of eHealth among GPII research project, carried out by the European Commission Directorate General for Communication Networks, Content & Technology. Open access microdata are available from the European Commission Digital Economy & Society Webpage.

Authors' Contributions

JTS participated in formulating the research question, study design, literature review, data analysis and statistical modeling, interpretation of the findings, and drafting the manuscript. ÁDC participated in data analysis and statistical modeling. ISR participated in the design and data analysis. FSR contributed to formulating the research question, study design, literature review, interpretation of results, and drafting the manuscript; he is the guarantor of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Statistical information based on General Practitioners II (GPII) survey.

[PDF File (Adobe PDF File), 26KB - jmir_v20i10e279_app1.pdf]

Multimedia Appendix 2

Benchmarking Deployment of eHealth among General Practitioners II (GPII) Questionnaire.

[PDF File (Adobe PDF File), 101KB - jmir_v20i10e279_app2.pdf]

Multimedia Appendix 3

Information and communication technology (ICT) usage by European general practitioners descriptive statistics 2012-2013.

[PDF File (Adobe PDF File), 34KB - jmir_v20i10e279_app3.pdf]

Multimedia Appendix 4

Barriers to eHealth usage by European general practitioners descriptive statistics 2012-2013.

[PDF File (Adobe PDF File), 34KB - jmir_v20i10e279_app4.pdf]

Multimedia Appendix 5

Personal Health Records (PHR) usage by European general practitioners descriptive statistics 2012-2013.

[PDF File (Adobe PDF File), 32KB - jmir_v20i10e279_app5.pdf]

Multimedia Appendix 6

Telehealth usage by European general practitioners descriptive statistics 2012-2013.

[PDF File (Adobe PDF File), 31KB - jmir_v20i10e279_app6.pdf]

Multimedia Appendix 7

Health Information Exchanges (HIE) usage by European general practitioners descriptive statistics 2012-2013.

[PDF File (Adobe PDF File), 35KB - jmir_v20i10e279_app7.pdf]

Multimedia Appendix 8

Electronic Health Records_ Decisions Support System (EHR_DSS) usage by European general practitioners descriptive statistics 2012-2013.

[PDF File (Adobe PDF File), 32KB - jmir_v20i10e279_app8.pdf]

Multimedia Appendix 9

Electronic Health Records_ Data (EHR_DAT) usage by European general practitioners descriptive statistics 2012-2013.

[PDF File (Adobe PDF File), 36KB - jmir_v20i10e279_app9.pdf]

Multimedia Appendix 10

Outcomes of internal practices in eHealth usage by European general practitioners descriptive statistics 2012-2013.

[PDF File (Adobe PDF File), 35KB - jmir_v20i10e279_app10.pdf]

Multimedia Appendix 11

Outcomes of practices with patients in eHealth usage by European general practitioners descriptive statistics 2012-2013.

[PDF File (Adobe PDF File), 35KB - jmir_v20i10e279_app11.pdf]

Multimedia Appendix 12

Explanatory factors of some dimensions of European general practitioners' eHealth usage and eHealth usage outcomes in 2012-2013.

[PDF File (Adobe PDF File), 39KB - jmir_v20i10e279_app12.pdf]

Multimedia Appendix 13

eHealth usage by European general practitioners dimensions frequency statistics* 2012-2013.

[PDF File (Adobe PDF File), 32KB - jmir_v20i10e279_app13.pdf]

References

1. Dinesen B, Nonnecke B, Lindeman D, Toft E, Kidholm K, Jethwani K, et al. Personalized Telehealth in the Future: A Global Research Agenda. J Med Internet Res 2016 Mar 01;18(3):e53 [FREE Full text] [doi: 10.2196/jmir.5257] [Medline: 26932229]
2. European Commission. Brussels, Belgium; 2012. eHealth Action Plan 2012-2020 – innovative healthcare for the 21st century Internet URL: https://ec.europa.eu/health/sites/health/files/ehealth/docs/com_2012_736_en.pdf [WebCite Cache ID 72bQ31mRS]
3. Kuperman GJ. Health-information exchange: why are we doing it, and what are we doing? J Am Med Inform Assoc 2011;18(5):678-682 [FREE Full text] [doi: 10.1136/amiajnl-2010-000021] [Medline: 21676940]
4. Roehrs A, da Costa CA, Righi RDR, de Oliveira KSF. Personal Health Records: A Systematic Literature Review. J Med Internet Res 2017 Dec 06;19(1):e13 [FREE Full text] [doi: 10.2196/jmir.5876] [Medline: 28062391]
5. European Commission.: Brussels, Belgium; 2014. Digital Agenda for Europe: A Europe 2020 Initiative [Internet] URL: http://ec.europa.eu/digital-agenda/en
6. Huygens M, Vermeulen J, Friele R, van Schayck OC, de Jong JD, de Witte LP. Internet Services for Communicating With the General Practice: Barely Noticed and Used by Patients. Interact J Med Res 2015 Nov 24;4(4):e21 [FREE Full text] [doi: 10.2196/ijmr.4245] [Medline: 26601596]
7. Peeters J, Krijgsman J, Brabers A, 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]
8. Kidholm K, Stafylas P, Kotzeva A, Duedal Pedersen C, Dafoulas G, Scharf I, Giannaokopoulos, et al. REgioNs of Europe WorkINg toGether for HEALTH Final Report - Public. 2014. URL: http://www.renewinghealth.eu/documents/28946/555381f3-9686-4955-8547-76b58be34a04 [WebCite Cache ID 72bSBCF8L]
9. Kidholm K, Stafylas P, Kotzeva A, Duedal Pedersen C, Dafoulas G, Scharf I. REgioNs of Europe WorkINg toGether for HEALTH Final Report - Public. 2014. URL: http://www.renewinghealth.eu/documents/28946/555381f3-9686-4955-8547-76b58be34a04 [WebCite Cache ID 72bSBCF8L]
10. Riazul Islam SM, Daehan Kwak, Humaun Kabir M, Hossain M, Kyung-Sup Kwak. The Internet of Things for Health Care: A Comprehensive Survey. IEEE Access 2015;3:678-708 [FREE Full text] [doi: 10.1109/ACCESS.2015.2437951]
11. Gubbi J, Buyya R, Marusic S, Palaniswami M. Internet of Things (IoT): A vision, architectural elements, and future directions. Future Generation Computer Systems 2013 Sep;29(7):1645-1660 [FREE Full text] [doi: 10.1016/j.future.2013.01.010]
12. Li S, Xu L, Zhao S. The internet of things: a survey. Inf Syst Front 2014 Apr 26;17(2):243-259 [FREE Full text] [doi: 10.1007/s10796-014-9492-7]
13. Jones M, Koziel C, Larsen D, Berry P, Kubatka-Willms E. Progress in the Enhanced Use of Electronic Medical Records: Data From the Ontario Experience. JMIR Med Inform 2017 Feb 22;5(1):e5 [FREE Full text] [doi: 10.2196/medinform.6928] [Medline: 28228372]
14. Ghazisaeedi M, Mohammadzadeh N, Safdari R. Electronic Health Record (EHR) As a Vehicle for Successful Health Care Best Practice. Med Arch 2014 Dec;68(6):419-421 [FREE Full text] [doi: 10.5455/medarh.2014.68.419-421] [Medline: 25648601]
15. Green LA, Potworowski G, Day A, May-Gentile R, Vibbert D, Maki B, et al. Sustaining. Ann Fam Med 2015;13(1):17-22 [FREE Full text] [doi: 10.1370/afm.1740] [Medline: 25583887]
16. Shea C, Malone R, Weinberger M, Reiter K, Thornhill J, Lord J, et al. Assessing organizational capacity for achieving meaningful use of electronic health records. Health Care Manage Rev 2014;39(2):124-133 [FREE Full text] [doi: 10.1097/HMR.0b013e3182860937] [Medline: 23380882]

17. Kruse C, Argueta D, 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]
18. Hogan T, Luger T, Volkman J, Rocheleau M, Mueller N, Barker A, et al. Patient Centeredness in Electronic Communication: Evaluation of Patient-to-Health Care Team Secure Messaging. J Med Internet Res 2018 Mar 08;20(3):e82 [FREE Full text] [doi: 10.2196/jmir.8801] [Medline: 29519774]
19. Ammenwerth E, Schnell-Inderst P, Hoerbst A. The impact of electronic patient portals on patient care: a systematic review of controlled trials. J Med Internet Res 2012 Nov 26;14(6):e162 [FREE Full text] [doi: 10.2196/jmir.2238] [Medline: 23183044]
20. Veroff D, Marr A, Wennberg DE. Enhanced support for shared decision making reduced costs of care for patients with preference-sensitive conditions. Health Aff (Millwood) 2013 Feb;32(2):285-293 [FREE Full text] [doi: 10.1377/hlthaff.2011.0941] [Medline: 23381521]
21. Parsi K, Chambers C, Armstrong A. Cost-effectiveness analysis of a patient-centered care model for management of psoriasis. J Am Acad Dermatol 2012 Apr;66(4):563-570 [FREE Full text] [doi: 10.1016/j.jaad.2011.02.022] [Medline: 21835497]
22. Torrent-Sellens J, Díaz-Chao Á, Soler-Ramos I, Saigí-Rubió F. Modelling and Predicting eHealth Usage in Europe: A Multidimensional Approach From an Online Survey of 13,000 European Union Internet Users. J Med Internet Res 2016 Dec 22;18(7):e188 [FREE Full text] [doi: 10.2196/jmir.5605] [Medline: 27450189]
23. Codagnone C, Lupiañez-Villanueva F. Benchmarking Deployment of eHealth among General Practitioners - Final Report. In: European Union. Luxembourg. Publications Office of the European Union: European Commission. Directorate-General of Communications Networks. Content & Technology; 2013:114.
24. 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]
25. Marcoulides G, Moustaki I. Latent variable and latent structure models. New York: Routledge; 2012:-296.
26. Saigí-Rubió F, Torrent-Sellens J, Jiménez-Zarco A. Drivers of telemedicine use: comparative evidence from samples of Spanish, Colombian and Bolivian physicians. Implement Sci 2014 Oct 08;9:128 [FREE Full text] [doi: 10.1186/s13012-014-0128-6] [Medline: 25293651]
27. Carey M, Noble N, Mansfield E, Waller A, Henskens F, Sanson-Fisher R. The Role of eHealth in Optimizing Preventive Care in the Primary Care Setting. J Med Internet Res 2015 May 22;17(5):e126 [FREE Full text] [doi: 10.2196/jmir.3817] [Medline: 26001983]
28. Lacasta Tintorer D, Manresa Domínguez JM, Pujol-Rivera E, Flayeh Beneyto S, Mundet Tuduri X, Saigí-Rubió F. Keys to success of a community of clinical practice in primary care: a qualitative evaluation of the ECOPIH project. BMC Fam Pract 2018 May 09;19(1):56 [FREE Full text] [doi: 10.1186/s12875-018-0739-0] [Medline: 29743030]
29. Gagnon M, Nsangou ÉR, Payne-Gagnon J, Grenier S, Sicotte C. Barriers and facilitators to implementing electronic prescription: a systematic review of user groups' perceptions. J Am Med Inform Assoc 2014;21(3):535-541 [FREE Full text] [doi: 10.1136/amiajnl-2013-002203] [Medline: 24130232]
30. Steele Gray C, Gill A, Khan A, Hans P, Kuluski K, Cott C. The Electronic Patient Reported Outcome Tool: Testing Usability and Feasibility of a Mobile App and Portal to Support Care for Patients With Complex Chronic Disease and Disability in Primary Care Settings. JMIR Mhealth Uhealth 2016 Jun 02;4(2):e58 [FREE Full text] [doi: 10.2196/mhealth.5331] [Medline: 27256035]
31. Granja C, Janssen W, Johansen M. Factors Determining the Success and Failure of eHealth Interventions: Systematic Review of the Literature. J Med Internet Res 2018 May 01;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
32. Bashshur R, Shannon G, Krupinski E, Grigsby J. Sustaining and realizing the promise of telemedicine. Telemed J E Health 2013 May;19(5):339-345 [FREE Full text] [doi: 10.1089/tmj.2012.0282] [Medline: 23289907]
33. Praveen D, Patel A, Raghu A, Clifford G, Maulik P, Mohammad Abdul A, et al. SMARTHealth India: Development and Field Evaluation of a Mobile Clinical Decision Support System for Cardiovascular Diseases in Rural India. JMIR Mhealth Uhealth 2014 Dec 08;2(4):e54 [FREE Full text] [doi: 10.2196/mhealth.3568] [Medline: 25487047]
34. Saigi-Rubió F, Jiménez-Zarco A, Torrent-Sellens J. Determinants of the intention to use telemedicine: evidence from primary care physicians. Int J Technol Assess Health Care 2016 Jul 29;32(1-2):29-36. [doi: 10.1017/S0266462316000015]
35. Díaz-Chao A, Torrent-Sellens J, Lacasta-Tintorer D, Saigí-Rubió F. Improving Integrated Care: Modelling the performance of an online community of practice. Int J Integr Care Internet 2014:1-10 [FREE Full text] [doi: 10.5334/ijic.1200]
36. Atun R. Copenhagen: WHO Regional Office for Europe (Health Evidence Network); 2004. What are the advantages and disadvantages of restructuring a health care system to be more focused on primary care services? Internet URL: http://www.euro.who.int/__data/assets/pdf_file/0004/74704/E82997.pdf
37. Macinko J, Starfield B, Shi L. The Contribution of Primary Care Systems to Health Outcomes within Organization for Economic Cooperation and Development (OECD) Countries, 1970-1998. Health Serv Res 2003 Jun;38(3):831-865 [FREE Full text] [doi: 10.1111/1475-6773.00149] [Medline: 12822915]

38. Ford E, Hesse B, Huerta T. 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. Janols R, Lind T, Göransson B, Sandblad B. Evaluation of user adoption during three module deployments of region-wide electronic patient record systems. Int J Med Inform 2014 Jun;83(6):438-449 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.02.003] [Medline: 24630924]
40. 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]
41. Goldstein DH, Phelan R, Wilson R, Ross-White A, VanDenKerkhof EG, Penning JP, et al. Brief review: Adoption of electronic medical records to enhance acute pain management. Can J Anaesth 2014 Feb;61(2):164-179. [doi: 10.1007/s12630-013-0069-6] [Medline: 24233770]
42. Jamoom E, Patel V, Furukawa M, King J. EHR adopters vs. non-adopters: Impacts of, barriers to, and federal initiatives for EHR adoption. Healthc (Amst) 2014 Mar;2(1):33-39 [FREE Full text] [doi: 10.1016/j.hjdsi.2013.12.004] [Medline: 26250087]
43. Ross J, Stevenson F, Lau R, Murray E. Exploring the challenges of implementing e-health: a protocol for an update of a systematic review of reviews. BMJ Open 2015 Apr 13;5(4):e006773 [FREE Full text] [doi: 10.1136/bmjopen-2014-006773] [Medline: 25869685]
44. Nazi K. 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 Apr 04;15(4):e70 [FREE Full text] [doi: 10.2196/jmir.2443] [Medline: 23557596]
45. Gagnon M, Desmartis M, Labrecque M, Car J, Pagliari C, Pluye P, et al. Systematic review of factors influencing the adoption of information and communication technologies by healthcare professionals. J Med Syst 2012 Feb;36(1):241-277 [FREE Full text] [doi: 10.1007/s10916-010-9473-4] [Medline: 20703721]
46. Embi P, Weir C, Efthimiadis E, Thielke S, Hedeen A, Hammond K. Computerized provider documentation: findings and implications of a multisite study of clinicians and administrators. J Am Med Inform Assoc 2013;20(4):718-726 [FREE Full text] [doi: 10.1136/amiajnl-2012-000946] [Medline: 23355462]
47. Hao W, Hsu Y, Chen K, Li H, Iqbal U, Nguyen PA, et al. LabPush: a pilot study of providing remote clinics with laboratory results via short message service (SMS) in Swaziland, Africa - a qualitative study. Comput Methods Programs Biomed 2015 Jan;118(1):77-83 [FREE Full text] [doi: 10.1016/j.cmpb.2014.10.005] [Medline: 25453385]

Abbreviations

BARRIERS: barriers to eHealth usage
CFI: comparative fit index
eHealth: electronic health
EHR: electronic health record
EHR_DSS: electronic health records_decision support system
EHR_DAT: electronic health records_data
GPII: Benchmarking Deployment of eHealth among General Practitioners Phase 2
HIE: health information exchange
HIT: health information technology
ICT: information and communication technology
ICTUS: information and communication technology usage
IFI: incremental fit index
IP: internal practices
NFI: normed fit index
OR: odds ratio
OUTINTPRA: eHealth usage outcomes in internal practices
OUTPRAPAT: eHealth usage outcomes in practices with patients
PHR: personal health record
PP: practices with patients
RFI: relative fit index
RMSEA: root mean square error of approximation
SEM: structural equation modeling
THEALTH: telehealth
TLI: Tucker-Lewis index

Edited by G Eysenbach; submitted 24.10.17; peer-reviewed by J Aerts, KC Wong; comments to author 27.02.18; revised version received 04.06.18; accepted 10.07.18; published 22.10.18.

Please cite as:
Torrent-Sellens J, Díaz-Chao Á, Soler-Ramos I, Saigí-Rubió F
Modeling and Predicting Outcomes of eHealth Usage by European Physicians: Multidimensional Approach from a Survey of 9196 General Practitioners
J Med Internet Res 2018;20(10):e279
URL: http://www.jmir.org/2018/10/e279/
doi:10.2196/jmir.9253
PMID:30348628

Original Paper

Oncology Patient Interest in the Use of New Technologies to Manage Their Disease: Cross-Sectional Survey

Roberto Collado-Borrell[1], PharmD; Vicente Escudero-Vilaplana[1], PharmD, PhD; Antonio Calles[1], MD; Estela Garcia-Martin[1], PharmD; Belen Marzal-Alfaro[1], PharmD; Eva Gonzalez-Haba[1], PharmD, PhD; Ana Herranz-Alonso[1], PharmD, PhD; Maria Sanjurjo-Saez[1], PharmD

Hospital General Universitario Gregorio Marañón, Madrid, Spain

Corresponding Author:
Vicente Escudero-Vilaplana, PharmD, PhD
Hospital General Universitario Gregorio Marañón
Calle de Dr Esquerdo 46
Madrid, 28007
Spain
Phone: 34 915867714 ext 477714
Email: vicente.escudero@salud.madrid.org

Abstract

Background: Information and communication technologies (ICTs) in oncology can revolutionize the medical care of cancer patients. ICTs can promote patients’ empowerment and real-time disease monitoring. There is limited information about the impact of ICTs in cancer patients or their level of interest in using these tools for greater management of their condition.

Objective: This study aimed to understand the ICT usage profile in hematology-oncology patients to identify their needs and determine their level of interest in these technologies as a means of managing their disease.

Methods: A 28-item questionnaire was drawn up by a multidisciplinary team including pharmacists and oncologists. The questions were organized into 3 blocks, which were as follows: block A—sociodemographic characteristics; block B—use of ICTs when searching for health-related information; and block C—usage preferences for health apps. Hematology-oncology patients receiving treatment between May and July 2017 were included. A paper copy of the questionnaire was handed over to patients in either the day hospital or the pharmaceutical care consultancy in pharmacy services.

Results: A total of 650 questionnaires were handed out, with a participation of 94.0% (611/650). Patient sociodemographic characteristics were as follows*:* mean age was 57.8 years (age range: 19-91). Of 611 participants, 40.7% (249/611) had a university education, and 45.1% (276/611) of participants reported their overall state of health to be good. Results from use of ICTs when searching for health-related information were as follows: 87.1% (532/611) of participants were interested in being informed about health-related matters. Of all participants, 75.5% (532/611) sought information from health professionals and 61.3% (375/611) on the internet. Before going to their doctor’s appointment, 21.8% (133/611) of patients looked up information about their disease or treatment on the internet. This access to the internet rose to 50.9% (311/611) after their first medical appointment with their oncologist. Usage preferences for health apps were as follows: 82.7% (505/611) had a smartphone, whereas 20.3% (124/611) had a health app installed. Overall, 81.5% (498/611) would use an app if their health professional recommended it to them, but 39.6% (242/611) were not willing to pay for it.

Conclusions: The hematology-oncology patients showed a great deal of interest in searching for health-related information by means of ICTs, especially using smartphones and apps. The issues that drew the most interest in terms of apps were appointment management, advice on disease management, and communication with health professionals. Free access to these features and the recommendation by a health professional are important factors when it comes to their use. Therefore, the health care provider is a key element in the recommendation of ICTs, providing their knowledge and experience concerning their correct usage.

(J Med Internet Res 2018;20(10):e11006) doi:10.2196/11006

KEYWORDS

eHealth; app; information and communication technologies; oncology patient

Introduction

Population aging, improvements in public health care, and therapeutic advances have given rise to an epidemiological pattern that is mainly represented by chronic diseases. This new patient profile introduces additional requirements in terms of patient care (as it corresponds to more detailed health information) and greater opportunities for communication and even novel telecare services. In this regard, information and communication technologies (ICTs) are bringing about a revolution in the management of chronic patients, acting as a source of information on one hand, while providing real-time monitoring for health professionals and two-way communication between both parties on the other [1-3]. The report *Los Ciudadanos ante la e-Sanidad* (Public Opinion on eHealth Care) by the Spanish National Observatory of Telecommunications and Information Society (ONTSI) estimated that 70.4% of the Spanish population owns a smartphone, and 74.4% of internet users refer to the internet for health-related information [4].

The introduction of mobile health has meant the classic features of ICTs (content, chats, and contacts) could be focused toward providing services for patients. The use of wireless and portable devices (smartphones, tablets, wearables, and smartwatches) facilitates the collection, processing, and transmission of health information of patient through mobile phone apps and remote monitoring devices. All these features can present several advantages such as developing the active role of patients in their own health care and affording health professionals' real-time access to patient clinical information [5-8].

Oncology is one of the medical fields that could benefit the most from these new tools [3,6]. Although cancer is associated with a high rate of morbidity and mortality, ongoing advances are producing higher survival rates, converting some of these diseases from fatal to chronic conditions [9]. In addition, patients are increasingly resorting to apps in search of information given the complexity of treatment and the numerous unknowns they must confront [3,10,11].

Currently, a wide range of health apps are available, particularly ones designed especially for oncology patients [3,12]. These apps have several features and functions, from the simplest intended for accessing information, to the more complex apps that can help guide the diagnosis or perform patient follow-up [3]. However, many of the apps are developed without taking into account preferences and/or needs of a patient [10]. In fact, only 40% of those available are considered to have any value for addressing health issues, and just 2% provide two-way communication between the patient and health professional [12]. Many of these apps also suffer from a lack of scientific validity and reliability. A review conducted in November 2014 found that only 48.8% of apps designed for cancer patients had been developed by health professionals [3]. As these tools are not being regulated by the relevant institutions, the main onus falls on health professionals to ensure the validity of the apps they recommend or prescribe to their patients [2,3,10]. The first step when prescribing a quality ICT is awareness of the true impact these technologies are having on lives of patients.

Therefore, our aim was to understand the usage profile of information and communication technologies in oncology patients to identify their needs and determine their degree of interest in these technologies as a means of managing their disease.

Methods

A 28-item questionnaire was designed by a multidisciplinary team, including pharmacists and oncologists, with experience of ICTs applied to the field of oncology. The questions were organized into the following 3 blocks: block A—sociodemographic characteristics of a patient (questions 1-9), block B—use of ICTs when searching for health-related information (questions 10-21), and block C—usage preferences for health apps (questions 22-28). The design of the questionnaire considered the model form created by ONTSI [4] and incorporated advice from a management consultancy firm specializing in process design, implementation, and improvement. With respect to the types of questions, the questionnaire comprised dichotomous (questions 2, 3, 6, 12, 18, 20, 22, and 27) with true or false response options or polytomous (questions 4-5, 7-10, 11, 13-17, 19, 21, and 23-26) categorical variables. A total of 10 patients completed a draft paper questionnaire to validate the fact that it was well understood.

Hematology-oncology patients in antineoplastic therapy (chemotherapy, immunotherapy, oral targeted therapies, and hormone therapy) and/or support therapy were included. Patients younger than 18 years and those who would have had trouble understanding the questions because of language or cultural barriers were excluded. Participation was voluntary and anonymous. The inclusion period ran from May to July 2017. A paper copy of the questionnaire was handed over to every patient in the day hospital (administration or waiting rooms) or the pharmaceutical care consultancy in the outpatient pharmacy services area. Alternatively, patients could take the questionnaire home and return it filled out at their next appointment.

The questionnaire was approved by the clinical research ethics committee of the hospital, project code EHPHDO2017, and conducted in accordance with the ethical principles laid down by the Declaration of Helsinki. It did not include any information about the personal data of patients to ensure data confidentiality.

Data were analyzed using the software SPSS version 21.0 (IBM Corp) for Windows. The variables were analyzed using descriptive statistics. Normality was analyzed by means of the Kolmogorov-Smirnov test. Numeric variables were compared with Student t test or the Mann-Whitney test. The association between qualitative variables was studied using Pearson chi-square test or Fisher exact test. Missing data were handled with simple imputation method. The corresponding measures of association and risk were calculated along with their CIs. Results with a value of $P<.05$ were considered statistically significant.

Results

Patients' Sociodemographic Characteristics

A total of 650 questionnaires were handed out, with a participation of 94% (we finally analyzed 611 questionnaires). Sociodemographic characteristics of patients are shown in Table 1. The mean age of patients was 57.8 years (range: 19-91). Of all participants, 61.9% (378/611) were women, 89.5% (547/611) lived with other people, and 40.7% (249/611) had a university education. Overall, 29.2% (179/611) of cases corresponded to hematologic tumors. Among nonhematologic tumors, the most frequent were breast (124/611, 20.3%), colorectal (113/611, 18.5%), and lung (60/611, 9.8%). Over half of the patients considered their state of health was either good (45.1%, 276/611) or very good (7.8%, 48/611).

Use of Information and Communication Technologies When Searching for Health-Related Information

Figure 1 shows the frequency with which patients used different ICTs to search for information on the internet. With regard to health-related information, 87.1% (532/611) of patients were interested in staying informed about health issues. Of all subjects, 75.5% (462/611) sought information from health professionals and 61.3% (375/611) on the internet. Of these, 71.2% (436/611) searched through Google and 21.9% (134/611) by means of social networks (YouTube, Twitter, and Facebook, among others), 20.3% (124/611) used medical societies, and 20.1% (123/611) sought information via patient-advocacy groups.

In response to the question concerning the reasons why patients looked for health information, 49.2% (301/611) replied they wanted to obtain information about disease prevention, healthy lifestyles, and health care, and 47.2% (289/611) were seeking information about the treatment their doctor had prescribed for them. We observed statistically significant differences depending on the level of education, age, sex, and tumor type. First, patients with secondary or university education and younger patients searched for more information on health ($P<.001$). Women sought more information regarding alternative medicines ($P<.001$) and how to get in touch with other people with similar health problems ($P=.04$). Patients with breast cancer were more likely to seek information about their treatment ($P=.03$) and alternative medicines ($P<.001$) than those with other types of tumor.

As for the question regarding whether the information they found on the internet was easy to understand, 10.8% (66/611) of patients replied "always." A higher proportion of young patients and those with secondary or university education responded "always" than those in other demographic groups ($P<.001$). For the question as to whether participants trusted the information they found on the internet, overall 21.1% (129/611) responded "no," but this was a statistically more frequent reply among older patients ($P<.001$).

Before going to their doctor's appointment, 21.8% (133/611) of patients turned to the internet for information about their disease and/or treatment, an action that was more common in younger participants ($P<.001$). After their appointment, 50.9% (311/611) of participants referred to the internet, again this was more typical among younger patients, as well as those with secondary or university education ($P<.001$). The rest of the responses to questions covering the use of ICTs when looking for information on health issues are detailed in Table 2.

Usage Preferences for Health Apps

Table 3 shows the results for preferences regarding the use of health apps. Of all participants surveyed, 82.7% (505/611) had a smartphone and 20.3% (124/611) had a health app. With respect to mobile phone use, 56.7% (347/611) of participants used theirs to access the internet and 42.8% (262/611) to use apps, among other functions. Of the variables analyzed here, we found that younger patients used apps to obtain information about their illness and its treatment ($P=.01$) and to manage their medical appointments and record and monitor their symptoms ($P<.01$) with a greater frequency than the other sociodemographic groups. Furthermore, patients with secondary or university education were more likely to use apps for disease prevention, health problems, and improving their lifestyle ($P=.004$), as well as for managing their appointments with the health care center or health professional ($P=.03$) than patients with different levels of education.

Figure 2 presents the reasons why our patients used health apps and their usage preferences. We observed that patients with secondary and/or university education were statistically more inclined than other the patient groups to use apps to manage medical appointments ($P<.01$), record and monitor their medication ($P<.01$), get in contact with health professionals ($P<.01$), and facilitate their remote monitoring ($P=.02$).

Table 1. Sociodemographic characteristics of patients (N=611).

Demographic characteristics	Patients, n (%)
Sex	
Male	233 (38.1)
Female	378 (61.9)
Age in years	
<55	256 (41.9)
56-65	179 (29.3)
>65	176 (28.8)
Lives alone?	
Yes	61 (10.0)
No	547 (89.5)
No response	3 (0.5)
Makes use of a telehomecare system (social nature service that facilitates daily life activities of dependents using technology)?	
No	576 (94.3)
Yes	35 (5.7)
Level of education	
No education or incomplete primary education	32 (5.2)
Primary education	110 (18.0)
Secondary education	220 (35.9)
University education	249 (40.7)
Some degree of dependence?	
No	528 (86.4)
Needs help once a day	29 (4.7)
Needs help several times a day	39 (6.4)
Totally dependent	14 (2.3)
No response	1 (0.2)
Type of cancer	
Bladder	9 (1.5)
Breast	124 (20.3)
Cervical	8 (1.3)
Colorectal	113 (18.5)
Esophageal	13 (2.1)
Head and neck	16 (2.6)
Hematological	179 (29.2)
Kidney	16 (2.6)
Lung	60 (9.8)
Ovarian	31 (5.1)
Pancreatic	28 (4.6)
Prostate	12 (2.0)
Sarcoma	16 (2.6)
Skin or melanoma	8 (1.3)
Stomach	11 (1.8)
Testicular	6 (1.0)

Demographic characteristics	Patients, n (%)
How would you rate your overall health?	
Very good	48 (7.9)
Good	276 (45.2)
Average	226 (37.0)
Poor	50 (8.2)
Very poor	8 (1.3)
No response	3 (0.2)

Figure 1. Frequency with which patients used different information and communication technologies to search for information on the internet.

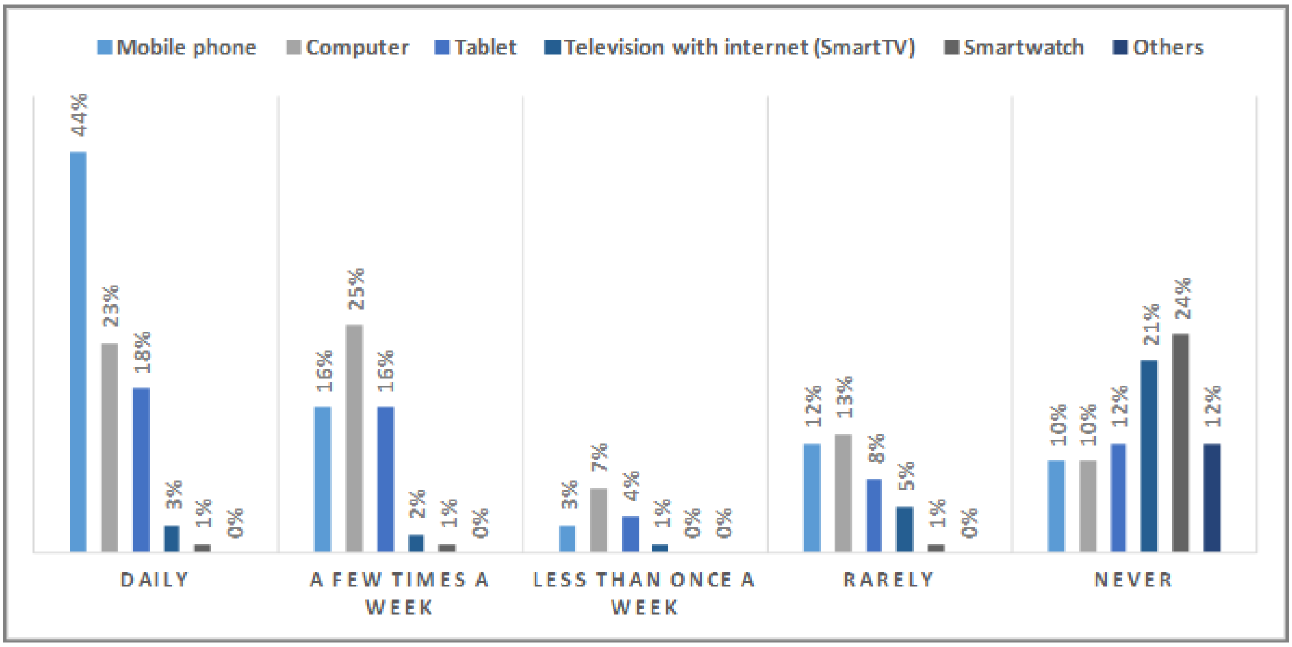


Table 2. Frequency of using new technologies to consult internet for oncohematological patients (N=611).

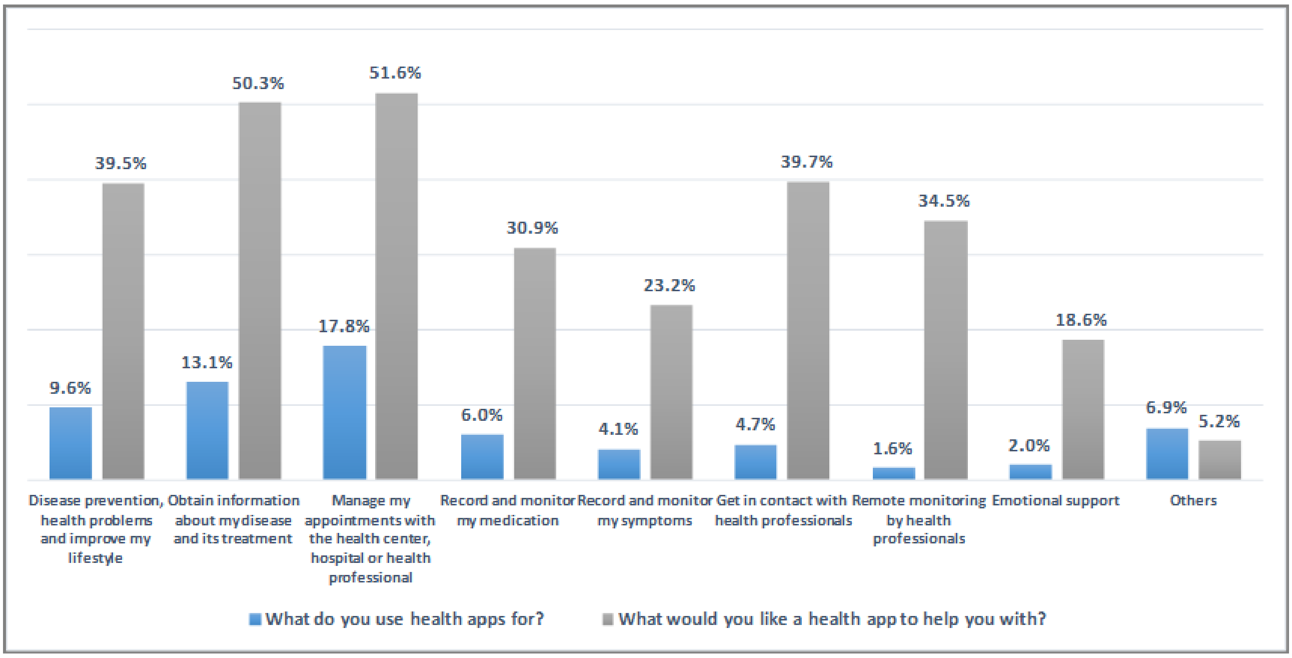
Use of information and communication technologies	Patients, n (%)
Which of the following terms have you heard before? (multiple answers are allowed)	
Mobile app	413 (67.5)
Smartphone	404 (66.0)
information and communication technologies	179 (29.2)
I have not heard any of these terms before	142 (23.2)
eHealth	88 (14.4)
Mobile health or mHealth	66 (10.8)
Wearable	51 (8.3)
Which devices do you use to look for information on the internet? (multiple answers are allowed)	
Mobile phone	445 (72.7)
Desktop and/or laptop computer	416 (68.0)
Tablet	274 (44.8)
Television with internet (SmartTV)	58 (9.5)
Smartwatch	15 (2.5)
Others	3 (0.5)
I do not look for information on the internet	84 (13.7)
Are you interested in staying informed about health-related matters?	
Yes	532 (87.1)
No	60 (9.8)
No response	19 (3.1)
Where do you search for information on health? (multiple answers are allowed)	
Health professionals	462 (75.5)
Internet	375 (61.3)
People close to me (friends, relatives, and workmates)	160 (26.1)
Newspapers, magazines, and pamphlets	151 (24.7)
Mobile apps	62 (10.1)
Others	23 (3.8)
If you use the internet to search for health information, which types of website do you use? (multiple answers are allowed)	
Google	436 (71.2)
I do not search for medical information on the internet	129 (21.1)
Medical societies	124 (20.3)
Patient associations	123 (20.1)
YouTube	68 (11.1)
Facebook	45 (7.4)
Blogs	43 (7.0)
Others	21 (3.4)
Other social networks	11 (1.8)
Twitter	10 (1.6)
For what purposes do you search for health information? (multiple answers are allowed)	
Disease prevention, healthy lifestyles, and health care	301 (49.2)
To find information about the treatment prescribed by my doctor	289 (47.2)
To find symptoms and learn about potential diseases	145 (23.7)

Use of information and communication technologies	Patients, n (%)
To find information about medical centers or health professionals	139 (22.7)
To find information about alternative or complementary medicines (herbal products and acupuncture)	107 (17.5)
To get in contact with other people with health problems such as mine	71 (11.6)
Others	69 (11.3)
Is it easy to understand the health information you find on the internet?	
Usually	269 (44.0)
Sometimes	148 (24.2)
No response	77 (12.6)
Always	66 (10.8)
Never	51 (8.3)
Do you trust the health information you find on the internet?	
Yes	84 (13.7)
No	129 (21.1)
Depends on the website	336 (55.0)
No response	62 (10.1)
Do you look up information on the internet about your disease and/or treatment *before* going to your doctor's appointment?	
No	451 (73.8)
Yes	133 (21.8)
No response	27 (4.4)
Do you look up information on the internet about your disease and/or treatment *after* going to your doctor's appointment?	
Yes	143 (23.4)
Only if I have still got doubts about something	168 (27.5)
No	270 (44.2)
No response	30 (4.9)

Table 3. Usage preferences for health apps (N=611).

Usage preferences for health apps	n (%)
Is your mobile phone a smartphone?	
No	93 (15.2)
Yes	505 (82.7)
No response	13 (2.1)
Do you have health-related apps installed?	
No	478 (78.2)
Yes	124 (20.3)
No response	9 (1.5)
What do you use your mobile phone for? (multiple answers are allowed)	
Normal phone use (calls, messages, and photos or videos)	591 (96.6)
To access the internet	347 (56.7)
Schedule planner and alarms	305 (49.8)
To use apps	262 (42.8)
Social networks	233 (38.1)
How would you like to communicate with your health professional? (multiple answers are allowed)	
Telephone	470 (76.8)
Email	270 (44.1)
Mobile apps	250 (40.8)
Videoconference	116 (19.0)
Website	51 (8.3)
Blogs	20 (3.3)
Social networks	18 (2.9)
Would you use an app if your health professional recommended it?	
Yes	498 (81.5)
No	82 (13.4)
No response	31 (5.1)
Would you download a health-related app if you had to pay approximately €2.15? (**considered the average price of the app aimed at patients with cancer)**	
No	242 (39.6)
Probably	199 (32.6)
Yes	102 (16.7)
No response	68 (11.1)

Figure 2. Reasons why patients used a health app, and things they wish the app could help them with.



Discussion

Principal Findings and Comparison With Prior Work

To our knowledge, this is the first survey aimed to identify the needs and interests of hematology-oncology patients in the ICT as a means of managing their disease. In this survey conducted in 611 patients, we observed that most of the patients (505/611, 82.7%) owned a smartphone. Furthermore, almost half of patients use their mobile, which was the most popular device among the different options, to perform internet searches every day.

Use of Information and Communication Technologies When Searching for Health-Related Information

We have seen that health is a cause for concern for 86.9% (531/611) of our patients. Traditionally, patients have tried to physically obtain medical information from ordinary health professionals (doctors, pharmacists, and nurses) [2,4,13]. However, the expansion and generalization of the use of ICTs has implied a radical change in how we obtain information about health and manage diseases [3,4,5,12,14]. Despite the fact that professional health care is still the main resource to obtain information on the pathology, 61.3% (375/611) of our patients turn to the internet to perform a query; similar data were reported by LC et al [15].

In addition, in this study, we have noted that 21.8% (133/611) of patients look up facts about their disease before visiting the doctor, a figure that can rise to 50% after their appointment. These data are akin to those reported by ONTSI [4]. This highlights the growing interest shown by patients and could represent a problem because of the heterogeneous nature of the data. The internet, without the necessary knowledge, is not the most reliable source of health-related information [10,13]. Of all participants in our survey, 71.2% (436/611) used search engines such as Google, and 10.1% (62/611) employed apps to look for information. However, only 10.8% (66/611) always found the information easy to understand. In this regard, the health professional must assume a proactive role when recommending the most reliable websites or apps and adapt them to the understanding of a patient [3]. According to the ONTSI, 38.9% of patients would like their doctor to recommend devices or apps to help them manage their health [4], yet only 10% to 20% of health professionals recommend websites, devices, or apps to their patients [4,16].

Usage Preferences for Health Apps

If we analyze interests of our patients, more than 40% (244/611) would like to be able to communicate with their health professional by means of an app or via email, a figure close to the 37.3% detailed in the ONTSI report published in April 2016 [4]. At this time, only 20.3% (124/611) of patients have a health-related app although 81.5% (498/611) would use them if recommended by their health care provider. There must be a certain guarantee about the accuracy and quality of the information on the apps, so this does not become a barrier for their own prescription. In a review of apps focused on cancer patients, it was found that only 48.8% were developed with the participation of a health care organization, only half of the apps had been updated in the last year, while 30% had not been updated for over 2 years [3].

Despite the interest these new technologies generate, the cost is a key factor governing their use. Considering the average price of apps designed for cancer patients is €2.15 [3], half of all patients would not download them if they had to pay this amount. Probably, this is because there is a culture among Spanish patients and relatives that everything provided by hospitals is paid for by the social security system and they have never had to pay for anything.

Finally, it is worth highlighting that although the world of apps is evolving considerably, they do not actually cover all our patients' needs [17,18]. Obtaining information about a disease or its treatment would interest 50% of our patients, which is more than the 28.1% reported by Soler et al [19]; however, less than 20% used apps for this purpose. Moreover, remote monitoring by health professionals is currently one of the least used functions, although it has the greatest potential for use [20-23], as 1 in 3 patients would like to have an app with this feature. Providing a new means of communication through ICTs could offer patients the opportunity to get involved in their disease and assume more responsibility with regard to monitoring their health, thus promoting patient empowerment [5,6,10,24,25]. That is why an easy-to-use app would prove to be a complete tool for this type of patient [3,5,6,13]. To date, existing apps for cancer patients concentrate solely on isolated problems—they do not cover all the needs associated with hematology-oncology patients [3,21-24].

Limitations

The survey is a type of quantitative research to obtain a large amount of information that objectively can be extrapolated. However, there are other qualitative techniques such as interviews or discussion groups that allow us to know in-depth the perceptions of patients. This type of technique can complete the surveys so that the analysis of the results is much more complete. The results of this survey are being taken into account in a larger project that we are carrying out. The aim of this project was to improve the integral circuit of oncohematological patient care. The semistructured interview methodology with groups of patients is being used. Another limitation of the survey was the number of questions, which could complicate its degree of fulfillment. The questionnaire was therefore handed out at the oncology day hospital so that patients could complete it while they received chemotherapy. Participants could also take the questionnaire home with them and return it completed at their next appointment that resulted in a very low percentage of dropouts. On the other hand, as the survey included new terms that were unfamiliar to some participants, they may not have understood some of the questions while completing the questionnaire. To overcome this limitation, the survey was mainly conducted in the oncology day hospital or outpatient clinic where collaborating researchers were available to help resolve any doubts patients may have had. The questionnaire also included a telephone number patients could call to clear up any doubts or questions that may arise while completing it at home. Finally, questions were completed by patients in Spain in the day hospital or the pharmaceutical care consultancy at pharmacy services. Thus, part of the results may not be representative of other cancer patients in different health systems. However, because we have described exhaustively the working methodology and the complete survey has been published, this work could be reproduced at any center.

Conclusions

The results of the survey show that hematology-oncology patients are very interested in finding health-related information via ICTs, especially using smartphones and apps. These apps, which are presented as tools that cover these needs, are used as both a source of information and as a new communication channel between patients and health professionals. Free access to apps is still an important factor when it comes to promoting their use and reaching a greater number of users. The app features that draw the most interest are appointment management, advice on disease management, and personalized communication with health professionals. However, many of the apps available have not been validated or updated and do not offer any alternatives to medical care. Therefore, the health professional plays an essential role in recommending and validating these tools to ensure they are of high quality standards.

Acknowledgments

The authors would like to thank the team at Speak Talk Communicate SL for their help preparing the manuscript and Gesco Consultores for advice on the survey. This project was supported by iPharma (Pharmacy Innovation Center at the Hospital General Universitario Gregorio Marañón) through funding received from ERDF Project PI13/02056; aid co-financed by the European Regional Development Fund “ERDF, a way of making Europe.”

Conflicts of Interest

None declared.

References

1. García Cuyás F, de San Pedro M, Ledesma Castelltor A. Las TICs y la gestión de pacientes crónicos. Sociedad Española Informática Salud 2014;105:11-13 [FREE Full text]
2. Cepeda Díez JM, Meijome Sánchez XM, Santillán García A. Innovaciones en salud y tecnologías: las cosas claras. Rev enferm CyL 2012;4(1):28-32 [FREE Full text]
3. Collado-Borrell R, Escudero-Vilaplana V, Ribed-Sánchez A, Ibáñez-García S, Herranz-Alonso A, Sanjurjo-Sáez M. Smartphone applications for cancer patients; what we know about them? Farm Hosp 2016 Jan 01;40(1):25-35 [FREE Full text] [doi: 10.7399/fh.2016.40.1.8993] [Medline: 26882831]
4. Los CALES. Ontsi.red. Opiniones y expectativas de los ciudadanos sobre el uso y aplicación de las TIC en el ámbito sanitario URL: http://www.ontsi.red.es/ontsi/sites/ontsi/files/los_ciudadanos_ante_la_e-sanidad.pdf [WebCite Cache ID 6y5UGRGpy]

5. Sawesi S, Rashrash M, Phalakornkule K, Carpenter JS, Jones JF. The impact of information technology on patient engagement and health behavior change: a systematic review of the literature. JMIR Med Inform 2016;4(1):e1 [FREE Full text] [doi: 10.2196/medinform.4514] [Medline: 26795082]
6. Weaver A, Young AM, Rowntree J, Townsend N, Pearson S, Smith J, et al. Application of mobile phone technology for managing chemotherapy-associated side-effects. Ann Oncol 2007 Nov;18(11):1887-1892 [FREE Full text] [doi: 10.1093/annonc/mdm354] [Medline: 17921245]
7. Kruse CS, Beane A. Health information technology continues to show positive effect on medical outcomes: systematic review. J Med Internet Res 2018 Feb 05;20(2):e41 [FREE Full text] [doi: 10.2196/jmir.8793] [Medline: 29402759]
8. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017 Apr 19;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]
9. Phillips JL, Currow DC. Cancer as a chronic disease. Collegian 2010;17(2):47-50. [Medline: 20738055]
10. Bender JL, Yue RY, To MJ, Deacken L, Jadad AR. A lot of action, but not in the right direction: systematic review and content analysis of smartphone applications for the prevention, detection, and management of cancer. J Med Internet Res 2013;15(12):e287 [FREE Full text] [doi: 10.2196/jmir.2661] [Medline: 24366061]
11. Highland J, Clemens H. Docplayer. Mandated benefit review of SB 1070: An act relative to oral cancer therapy URL: http://docplayer.net/20040894-Mandated-benefit-review-of-sb-1070-an-act-to-relative-to-oral-cancer-therapy.html [accessed 2018-03-21] [WebCite Cache ID 6y5U24eIG]
12. Aitken M, Lyle J. Microsites.schott. Patient Adoption of mHealth: Use, Evidence and Remaining Barriers to Mainstream Acceptance URL: http://microsites.schott.com/d/studentchallenge/346d3edb-327f-42ad-b3b5-d859b8c4cf6f/1.0/ims_report_patient_adoption_of_mhealth.pdf [WebCite Cache ID 6y5U7geG9]
13. Mira JJ, Pérez-Jover V, Lorenzo S. [Navigating on the internet in search of health information: all that glitters is not gold] [Article in Spanish]. Aten Primaria 2004 Apr 30;33(7):391-399 [FREE Full text] [Medline: 15117635]
14. Kessel KA, Vogel MM, Kessel C, Bier H, Biedermann T, Friess H, et al. Mobile health in oncology: a patient survey about app-assisted cancer care. JMIR Mhealth Uhealth 2017 Jun 14;5(6):e81 [FREE Full text] [doi: 10.2196/mhealth.7689] [Medline: 28615159]
15. An LC, Wallner L, Kirch MA. Online social engagement by cancer patients: a clinic-based patient survey. JMIR Cancer 2016 Aug 19;2(2):e10 [FREE Full text] [doi: 10.2196/cancer.5785] [Medline: 28410186]
16. 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]
17. Robbins R, Krebs P, Jagannathan R, Jean-Louis G, Duncan DT. Health app use among US mobile phone users: analysis of trends by chronic disease status. JMIR Mhealth Uhealth 2017 Dec 19;5(12):e197 [FREE Full text] [doi: 10.2196/mhealth.7832] [Medline: 29258981]
18. Halwas N, Griebel L, Huebner J. eHealth literacy, internet and eHealth service usage: a survey among cancer patients and their relatives. J Cancer Res Clin Oncol 2017 Nov;143(11):2291-2299. [doi: 10.1007/s00432-017-2475-6] [Medline: 28699035]
19. Soler Montaner I, García Muñoz S, Sangrador Pelluz C, Soler Montaner B, Soler Company E. SEFAF. Valencia. Spain: Sociedad Española de Farmacoterapia y Atención Farmacéutica (SEFAF); 2014. Paciente al día, atención farmacéutica al día. Impacto de las TIC en las estrategias de mejora de la adherencia terapéutica URL: http://sefaf.com/wp-content/uploads/2017/07/Paciente-al-d%C3%ADa-Atenci%C3%B3n-Farmac%C3%A9utica-al-d%C3%ADa.pdf [accessed 2018-09-25] [WebCite Cache ID 72hJsCNFm]
20. Iqvia. The Growing Value of Digital Health: Evidence and Impact on Human Health and the Healthcare System URL: https://www.iqvia.com/institute/reports/the-growing-value-of-digital-health [WebCite Cache ID 6y5TyEi2t]
21. Voruganti T, Husain A, Grunfeld E, Webster F. Disruption or innovation? A qualitative descriptive study on the use of electronic patient-physician communication in patients with advanced cancer. Support Care Cancer 2018 Mar 04:2785-2792. [doi: 10.1007/s00520-018-4103-7] [Medline: 29502156]
22. Grünloh C, Myreteg G, Cajander Å, Rexhepi H. "Why Do They Need to Check Me?" Patient participation through eHealth and the doctor-patient relationship: qualitative study. J Med Internet Res 2018 Jan 15;20(1):e11 [FREE Full text] [doi: 10.2196/jmir.8444] [Medline: 29335237]
23. Kessel KA, Vogel MM, Schmidt-Graf F, Combs SE. Mobile apps in oncology: a survey on health care professionals' attitude toward telemedicine, mHealth, and oncological apps. J Med Internet Res 2016 Nov 24;18(11):e312 [FREE Full text] [doi: 10.2196/jmir.6399] [Medline: 27884810]
24. Ernsting C, Dombrowski SU, Oedekoven M, O Sullivan JL, Kanzler M, Kuhlmey A, et al. Using smartphones and health apps to change and manage health behaviors: a population-based survey. J Med Internet Res 2017 Apr 05;19(4):e101 [FREE Full text] [doi: 10.2196/jmir.6838] [Medline: 28381394]
25. Liu HY, Lee WC, Sun YC, Fen JJ, Chen TJ, Chou LF, et al. Hospital-owned apps in Taiwan: nationwide survey. JMIR Mhealth Uhealth 2018 Jan 16;6(1):e22 [FREE Full text] [doi: 10.2196/mhealth.8636] [Medline: 29339347]

Abbreviations

ICT: information and communication technology
ONTSI: Spanish National Observatory of Telecommunications and Information Society

---

Edited by G Eysenbach; submitted 09.05.18; peer-reviewed by J Ringwald, S Davis; comments to author 18.07.18; revised version received 20.07.18; accepted 23.07.18; published 23.10.18.

Please cite as:
Collado-Borrell R, Escudero-Vilaplana V, Calles A, Garcia-Martin E, Marzal-Alfaro B, Gonzalez-Haba E, Herranz-Alonso A, Sanjurjo-Saez M
Oncology Patient Interest in the Use of New Technologies to Manage Their Disease: Cross-Sectional Survey
J Med Internet Res 2018;20(10):e11006
URL: http://www.jmir.org/2018/10/e11006/
doi:10.2196/11006
PMID:30355554

Review

Skin Cancer Classification Using Convolutional Neural Networks: Systematic Review

Titus Josef Brinker[1,2], MD; Achim Hekler[1]; Jochen Sven Utikal[3,4], MD; Niels Grabe[5], PhD; Dirk Schadendorf[6], MD; Joachim Klode[6], MD; Carola Berking[7], MD; Theresa Steeb[7], MPH; Alexander H Enk[2], MD; Christof von Kalle[1], MD, PhD

[1]National Center for Tumor Diseases, Department of Translational Oncology, German Cancer Research Center, Heidelberg, Germany

[2]Department of Dermatology, University Hospital Heidelberg, University of Heidelberg, Heidelberg, Germany

[3]Skin Cancer Unit, German Cancer Research Center, Heidelberg, Germany

[4]Department of Dermatology, Venereology and Allergology, University Medical Center Mannheim, Ruprecht-Karl University of Heidelberg, Heidelberg, Germany

[5]Bioquant, Hamamatsu Tissue Imaging and Analysis Center, University of Heidelberg, Heidelberg, Germany

[6]Department of Dermatology, University Hospital of Essen, University of Duisburg-Essen, Essen, Germany

[7]Department of Dermatology, University Hospital Munich, Ludwig Maximilian University of Munich, Munich, Germany

Corresponding Author:
Titus Josef Brinker, MD
National Center for Tumor Diseases
Department of Translational Oncology
German Cancer Research Center
Im Neuenheimer Feld 460
Heidelberg, 69120
Germany
Phone: 49 15175084347
Email: titus.brinker@nct-heidelberg.de

Abstract

Background: State-of-the-art classifiers based on convolutional neural networks (CNNs) were shown to classify images of skin cancer on par with dermatologists and could enable lifesaving and fast diagnoses, even outside the hospital via installation of apps on mobile devices. To our knowledge, at present there is no review of the current work in this research area.

Objective: This study presents the first systematic review of the state-of-the-art research on classifying skin lesions with CNNs. We limit our review to skin lesion classifiers. In particular, methods that apply a CNN only for segmentation or for the classification of dermoscopic patterns are not considered here. Furthermore, this study discusses why the comparability of the presented procedures is very difficult and which challenges must be addressed in the future.

Methods: We searched the Google Scholar, PubMed, Medline, ScienceDirect, and Web of Science databases for systematic reviews and original research articles published in English. Only papers that reported sufficient scientific proceedings are included in this review.

Results: We found 13 papers that classified skin lesions using CNNs. In principle, classification methods can be differentiated according to three principles. Approaches that use a CNN already trained by means of another large dataset and then optimize its parameters to the classification of skin lesions are the most common ones used and they display the best performance with the currently available limited datasets.

Conclusions: CNNs display a high performance as state-of-the-art skin lesion classifiers. Unfortunately, it is difficult to compare different classification methods because some approaches use nonpublic datasets for training and/or testing, thereby making reproducibility difficult. Future publications should use publicly available benchmarks and fully disclose methods used for training to allow comparability.

(J Med Internet Res 2018;20(10):e11936) doi:10.2196/11936

KEYWORDS

skin cancer; convolutional neural networks; lesion classification; deep learning; melanoma classification; carcinoma classification

Introduction

In the past 10-year period, from 2008 to 2018, the annual number of melanoma cases has increased by 53%, partly due to increased UV exposure [1,2]. Although melanoma is one of the most lethal types of skin cancer, a fast diagnosis can lead to a very high chance of survival.

The first step in the diagnosis of a malignant lesion by a dermatologist is visual examination of the suspicious skin area. A correct diagnosis is important because of the similarities of some lesion types; moreover, the diagnostic accuracy correlates strongly with the professional experience of the physician [3]. Without additional technical support, dermatologists have a 65%-80% accuracy rate in melanoma diagnosis [4]. In suspicious cases, the visual inspection is supplemented with dermatoscopic images taken with a special high-resolution and magnifying camera. During the recording, the lighting is controlled and a filter is used to reduce reflections on the skin, thereby making deeper skin layers visible. With this technical support, the accuracy of skin lesion diagnosis can be increased by a further 49% [5]. The combination of visual inspection and dermatoscopic images ultimately results in an absolute melanoma detection accuracy of 75%-84% by dermatologists [6,7].

For some time, the problem of classifying skin lesions has also moved into the focus of the machine learning community. Automated lesion classification can both support physicians in their daily clinical routine and enable fast and cheap access to lifesaving diagnoses, even outside the hospital, through installation of apps on mobile devices [8,9]. Before 2016, research mostly followed the classical workflow of machine learning: preprocessing, segmentation, feature extraction, and classification [9-11]. However, a high level of application-specific expertise is required, particularly for feature extraction, and the selection of adequate features is very time-consuming. In addition, errors and the loss of information in the first processing steps have a very strong influence on the classification quality. For example, a poor segmentation result often leads to poor results in feature extraction and, consequently, low classification accuracy.

In 2016, a change occurred regarding the research of lesion classification techniques. An indication of this change can be found in the methods submitted to the 2016 International Symposium on Biomedical Imaging (ISBI) [12]. The 25 participating teams did not apply traditional standard machine learning methods; instead, they all employed a deep learning technique: convolutional neural networks (CNNs) [13].

This paper presents the first systematic review of the state-of-the-art research on classifying skin lesions using CNNs. The presented methods are categorized by whether a CNN is used exclusively as a feature extractor or whether it is applied for end-to-end-learning. The conclusion of this paper discusses why the comparability of the presented techniques is very difficult and which challenges must be addressed in the future.

Methods

Search Strategy

The Google Scholar, PubMed, Medline, ScienceDirect, and Web of Science databases were searched for systematic reviews and original research articles published in English. The search terms *convolutional neural networks*, *deep learning*, *skin cancer*, *lesions*, *melanoma*, and *carcinoma* were combined. Only papers that showed sufficient scientific proceedings are included in this review.

Study Selection

We limited our review to skin lesion classification methods. In particular, methods that apply a CNN only for lesion segmentation or for the classification of dermatoscopic patterns as in Demyanov et al [14] are not considered in this paper. Furthermore, only papers that show a sufficient scientific proceeding are included in this review. This latter criterion includes presenting the approaches in an understandable manner and discussing the results sufficiently. Works in which the origin of the performance was not plausible are not considered in this work, for example, in Carcagnì et al [15] or Dorj et al [16].

Convolutional Neural Networks

CNNs are neural networks with a specific architecture that have been shown to be very powerful in areas such as image recognition and classification [17]. CNNs have been demonstrated to identify faces, objects, and traffic signs better than humans and therefore can be found in robots and self-driving cars.

CNNs are a supervised learning method and are therefore trained using data labeled with the respective classes. Essentially, CNNs learn the relationship between the input objects and the class labels and comprise two components: the hidden layers in which the features are extracted and, at the end of the processing, the fully connected layers that are used for the actual classification task. Unlike regular neural networks, the hidden layers of a CNN have a specific architecture. In regular neural networks, each layer is formed by a set of neurons and one neuron of a layer is connected to each neuron of the preceding layer. The architecture of hidden layers in a CNN is slightly different. The neurons in a layer are not connected to all neurons of the preceding layer; rather, they are connected to only a small number of neurons. This restriction to local connections and additional pooling layers summarizing local neuron outputs into one value results in translation-invariant features. This results in a simpler training procedure and a lower model complexity.

Current Classifiers for Skin Lesions Based on Convolutional Neural Networks

In this section, the individual CNN methods used to classify skin lesions are presented. CNNs can be used to classify skin lesions in two fundamentally different ways. On the one hand, a CNN pretrained on another large dataset, such as ImageNet [18], can be applied as a feature extractor. In this case, classification is performed by another classifier, such as k-nearest neighbors, support vector machines, or artificial neural networks. On the other hand, a CNN can directly learn the relationship between the raw pixel data and the class labels through end-to-end learning. In contrast with the classical workflow typically applied in machine learning, feature extraction becomes an integral part of classification and is no longer considered as a separate, independent processing step. If the CNN is trained by end-to-end learning, the research can be additionally divided into two different approaches: learning the model from scratch or transfer learning. An overview of the presented CNN methods is shown in Figure 1.

A basic requirement for the successful training of deep CNN models is that sufficient training data labeled with the classes are available. Otherwise, there is a risk of overfitting the neural network and, as a consequence, an inadequate generalization property of the network for unknown input data. There is a very limited amount of data publicly available for the classification of skin lesions. Almost all published methods use datasets that contain far less than 1000 training data points per training class. In comparison, well-known CNN models for image classification, such as AlexNet [18], VGG [19], GoogLeNet [20], or ResNet [21], are trained via the large image database ImageNet and have over 1000 training images for each training class.

However, through the use of a specific training procedure called transfer learning, powerful CNN models with several million free parameters can also be employed for classification, even if only a small amount of data are available for training. In this case, the CNN is pretrained using a very large dataset, such as ImageNet; it is then used as an initialization of the CNN for the respective task. In particular, the last fully connected layer of the pretrained CNN model is modified according to the number of training classes in the actual classification task. There are then two options for the weights of the pretrained CNN: to fine-tune all layers of the CNN or to freeze some of the front layers because of overfitting problems and to fine-tune only some back layers of the network. The idea behind this technique is that that the front layers of a CNN contain more generic features (eg, edge or color-blob detectors) that are useful for many tasks, but the back layers of the CNN become increasingly specific to the details of the classes contained in the original dataset.

In the following discussion, statistical quantities to evaluate different classifiers are introduced. Next, methods that utilize the CNN as a feature extractor are presented. The last subsection provides an overview of the methods involved when using CNN for end-to-end-learning.

Performance Metrics for Classifiers

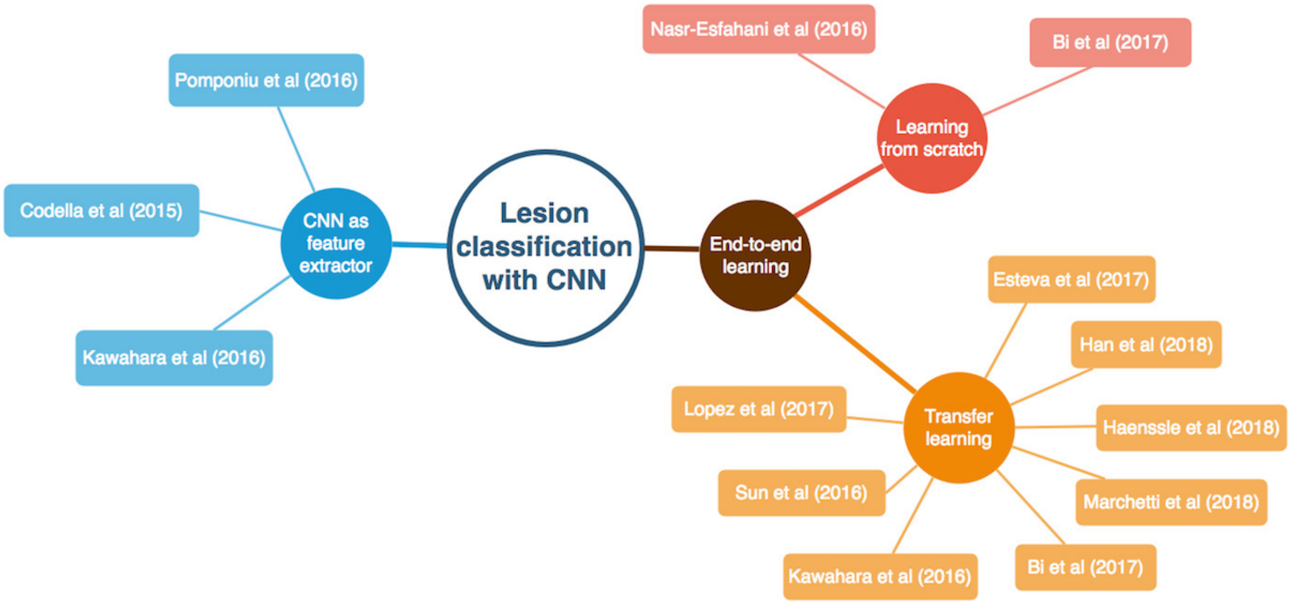
A classifier assigns each object to a class. This assignment is generally not perfect and objects may be assigned to the wrong class. To evaluate a classifier, the actual class of the objects must be known. To evaluate the classification quality, the class assigned by the classifier is compared with the actual class. This allows the objects to be divided into the following four subsets:

1. True positive (TP): the classifier correctly predicts the positive class.
2. True negative (TN): the classifier correctly predicts the negative class.
3. False positive (FP): the classifier incorrectly predicts the positive class.
4. False negative (FN): the classifier incorrectly predicts the negative class.

Based on the cardinality of these subsets, statistical quantities for the classifier can now be calculated. A common and widely used quantity is accuracy, which is only a reasonable measure if the different classes in the dataset are approximately equally distributed. Accuracy is calculated by (TP + TN)/(TP + TN + FP + FN). It specifies the percentage of objects that have been correctly classified.

Two other important metrics are sensitivity and specificity, which can be applied even if the different classes are not equally distributed. Sensitivity indicates the ratio of objects correctly classified as positive out of the total number of positive objects contained in the dataset and is calculated by TP/(TP + FN).

Figure 1. An overview of the presented convolutional neural networks (CNNs) methods and the corresponding categorization.



Specificity indicates the ratio of negative objects correctly classified as negative out of the total number of negative objects contained in the available dataset and is calculated by TN/(TN + FP).

The output of a binary classifier is interpreted as a probability distribution over the classes. Normally, objects with an output value greater than .5 are assigned to the positive class in a binary classifier and objects with an output value less than .5 are assigned to the negative class. An alternative approach is used based on the receiver operating characteristic (ROC). The threshold used for classification systematically varies between 0 and 1, and the sensitivity and specificity are determined for each selected threshold. The ROC curve is calculated by plotting the sensitivity against 1-specificity and can be used to evaluate the classifier. The further the ROC curve deviates from the diagonal, the better the classifier. A suitable overall measure for the curve is the area under the curve (AUC).

Results

Classifier That Utilizes the Convolutional Neural Network as a Feature Extractor

A CNN can be included in classification by removing the fully connected layers of a CNN that were pretrained with a large dataset. In skin lesion classification, pretraining is performed using ImageNet. Despite the nonmedical image domain, the learned features have sufficient quality for lesion classification [22].

Pomponiu et al [23] used only 399 images from a standard camera for the classification of melanomas versus benign nevi. First, data augmentation and preprocessing were performed. Subsequently, a pretrained AlexNet was applied for the extraction of representational features. The lesions were then classified with a k-nearest-neighbor classifier using cosine distance metrics. The algorithm was not tested with an independent test dataset; only a cross-validation was performed. The algorithm achieved a sensitivity of 92.1%, a specificity of 95.18%, and an accuracy of 93.64%. In addition to the nonexistent independent test dataset, it is also critical to note that the region of interest for each skin lesion must be manually annotated.

An AlexNet model for feature extraction was also applied by Codella et al [24]. In contrast to Gutman et al [12], however, a total of 2624 dermatoscopic images from the publicly available International Skin Imaging Collaboration (ISIC) database were used for the classification of melanoma versus nonmelanoma lesions or melanoma versus atypical nevi. In addition to the modified AlexNet outputs, the authors also used low-level handcrafted features and features from sparse coding, a deep residual network, and a convolutional U-network. Classification based on all of these features was then performed using a support vector machine. The authors reported an accuracy of 93.1%, a sensitivity of 94.9%, and a specificity of 92.8% for classifying melanoma versus nonmelanoma. In the more difficult discrimination between melanomas and atypical nevi, an accuracy of 73.9%, a sensitivity of 73.8%, and a specificity of 74.3% were reported. The authors also showed that the use of deep features results in a better performance compared to classifiers that only used low-level handcrafted features.

Kawahara et al [25] used a linear classifier to classify 10 different skin lesions. Feature extraction was also performed using an AlexNet whose last fully connected layer was replaced with a convolutional layer. This slightly modified AlexNet was tested using the public Dermofit Image Library, which contains 1300 clinical images of 10 skin lesions. An accuracy of 81.8% was achieved based on the entire dataset of 10 different types of skin lesions.

Skin Lesion Classifier Using End-to-End Learning

Convolutional Neural Network Model Training With Transfer Learning

Because publicly available datasets are limited, a common method of skin lesion classification involves transfer learning. Therefore, all such works pretrain a CNN via the ImageNet dataset; next, the weighting parameters of the CNN are fine-tuned to the actual classification problem.

Esteva et al [26] presented a landmark publication. For the first time, a CNN model was trained with a large amount of data, specifically 129,450 images, of which 3374 were obtained from dermatoscopic devices and represented 2032 different skin lesions. Two binary classification problems were considered: keratinocyte carcinomas versus benign seborrheic keratosis and malignant melanomas versus benign nevi. The last classification differentiation was performed for both clinical and dermatoscopic images. The authors used a GoogLeNet Inception v3 model for the classification, which was pretrained with the large image database ImageNet. The CNN model was then fine-tuned to classify skin lesions using transfer learning. A special property of this approach is the use of a novel tree-structured disease taxonomy in which the individual diseases form the leaves of the tree. The inner nodes group together individual diseases that are visually and clinically similar. The CNN does not have a two-dimensional vector as the output; instead, it reports a probability distribution with over 757 training classes. To determine the probabilities of a coarser lesion class (ie, an inner node at a higher level in the tree), the probabilities of the child nodes of this coarser lesion class are summed together. The authors show within the evaluation that a CNN that has been trained for finer classes has a better performance than a CNN that has been trained for the distinct classes that are of interest for the problem. The trained CNN was tested with test data that were fully biopsy-proofed and achieved an ROC AUC of .96 for carcinomas, an ROC AUC of .96 for melanomas, and an ROC AUC of .94 for melanomas classified exclusively with dermatoscopic images.

Haenssle et al [3] presented a very similar approach to Esteva et al [26]. A GoogLeNet Inception v3 model was adapted for skin lesion classification with transfer learning, whereby the weights were fine-tuned in all layers. The analysis was limited to dermatoscopic images of melanoma versus benign nevi and the AUC ROC achieved for this task was .86 (Esteva et al [26]: .94). The exact number of training data points was not provided and not all data had been proofed by a biopsy. However, the publication included the largest number of dermatologists to date (n=58) and was the first to indicate that additional clinical information improves both; sensitivity and specificity of dermatologists.

Han et al [27] are particularly noteworthy for their scientific transparency since they have made their computer algorithm publicly available for external testing. The team presented a classifier for 12 different skin diseases based on clinical images. They developed a ResNet model that was fine-tuned with 19,398 training images. With the publicly available Asan dataset, the CNN model achieved ROC AUCs for the diagnoses of basal cell carcinoma, squamous cell carcinoma, intraepithelial carcinoma, and melanoma of .96, .83, .82, and .96, respectively.

An ensemble of CNNs for the classification of melanomas versus nevi or lentigines is presented by Marchetti et al [13]. They implemented five methods to fuse all automated predictions from the 25 participating teams in the ISBI 2016 Challenge into a single classification result. For this purpose, they tested two nonlearning approaches and three machine learning methods. The fusion algorithms were trained with 279 dermatoscopic images from the ISBI 2016 Challenge dataset and were tested with 100 other dermatoscopic images from the same dataset. Based on average precision, greedy fusion was the best-performing ensemble method with a sensitivity of 58% and a specificity of 88%.

Another type of CNN ensemble was presented by Bi et al [28]. They considered the classification of melanomas versus seborrheic keratosis versus nevi using dermatoscopic images. They did not train multiple CNNs for the same classification problem; instead, three ResNets for different problems were trained: one for the original three-class problem and two binary classifiers (ie, melanoma versus both other lesion classes or seborrheic carcinoma versus both other lesion classes) by fine-tuning a pretrained CNN. The test utilized 150 dermatoscopic images and resulted in an ROC AUC of .854 for melanomas, an ROC AUC of .976 for seborrheic carcinomas, and an average ROC AUC over all classes of .915.

A special architecture of a CNN ensemble is presented by Kawahara et al [29]. The CNN was composed of multiple parts in which each part considers the same image at a different resolution. Next, an end layer combines the outputs from multiple resolutions into a single layer. The CNN identifies interactions across different image resolutions and the weighting parameters are completely optimized by end-to-end learning. The algorithm achieved an average classification accuracy of 79.5% in the public Dermofit Image Library.

Similar to Esteva et al [26], Sun et al [30] introduced a classifier that used 198 very finely defined training classes. In total, 6584 clinical images of the publicly available image archive DermQuest were used for training and testing and the performance of the CNN models CaffeNet and VGGNet were evaluated for this classification problem. The best average accuracy of 50.27% over all 198 classes was obtained using a pretrained VGGNet, which was optimized by fine-tuning the weighting parameters.

A modified VGGNet was also utilized by Lopez et al [31], where the classification of melanoma versus nevi or lentigines was addressed using dermatoscopic images. The authors compared the classification accuracy of a CNN trained from scratch, a pretrained CNN with transfer learning and frozen layers, and a pretrained CNN with transfer learning and fine-tuning of the weighting parameters. All three configurations were tested with 379 images from the ISBI 2016 Challenge dataset, and the last-mentioned configuration achieved the highest accuracy of 81.33%.

Convolutional Neural Network Model Training From Scratch

The previously introduced two-step approach by Bi et al [28] also falls under the category “learning from scratch” due to the method of training of the the ResNet model for the three-class classification of melanoma versus seborrheic keratosis versus nevus. Bi et al [28] used approximately 3600 dermatoscopic images from the ISBI 2017 Challenge dataset and additional images from the ISIC Archive to achieve the results reported.

In Nasr-Esfahani et al [32], a two-layer CNN was trained from scratch for the distinction of melanoma versus benign nevi based on clinical images. Only 136 images were used to train the model and the test dataset contained only 34 images. The images were all from the public image archive of the Department of Dermatology of the University Medical Center Groningen. The method achieved a sensitivity of 81%, a specificity of 80%, and an accuracy of 81%. However, the result should be viewed critically because the test dataset was very limited.

Discussion

Principal Findings

One issue with the comparison of skin lesion classification methods is that the considered problem formulations of the individual works differ, sometimes only slightly. This occurs not only for the considered training classes and the used data, but also for the presented statistical quantities. In addition, some works use nonpublic archives of skin clinics in addition to publicly accessible data archives [3,26]. This makes it even more difficult to reproduce the results. Since 2016, the ISIC Melanoma Project has attempted to improve this aspect by establishing a publicly accessible archive of dermatoscopic skin lesion images as a benchmark for education and research [12]. In addition, they announced an annual challenge in which a clearly defined problem must be addressed. It would be desirable if more work would compare itself with this benchmark to achieve a better ranking of the procedures in the state of research.

Another important challenge in this research area is the development of large public image archives with images as representative of the world population as possible [33]. The existing image archives mainly contain skin lesions from light-skinned people. The images in the ISIC database, for example, come mainly from the United States, Europe, and Australia. To achieve an accurate classification for dark-skinned people as well, the CNN must learn to abstract from the skin color. However, this can only occur if it observes enough pictures of dark-skinned people during the training.

An improvement in classification quality could be achieved by adding clinical data (eg, age, gender, race, skin type, and anatomic location) as inputs for the classifiers. This additional information is advantageous for the decision making of dermatologists, as Haenssle et al [3] show. Future work should take these aspects into account.

Conclusions

Unfortunately, it is difficult and in many times impossible to compare the performance of published classification results since many authors use nonpublic datasets for training and/or testing. Future publications should use publicly available benchmarks and fully disclose methods used for training to allow comparability.

Conflicts of Interest

None declared.

References

1. Nami N, Giannini E, Burroni M, Fimiani M, Rubegni P. Teledermatology: State-of-the-art and future perspectives. Expert Rev Dermatol 2014 Jan 10;7(1):1-3. [doi: 10.1586/edm.11.79]
2. Fabbrocini G, Triassi M, Mauriello MC, Torre G, Annunziata MC, De Vita V, et al. Epidemiology of skin cancer: Role of some environmental factors. Cancers (Basel) 2010 Nov 24;2(4):1980-1989 [FREE Full text] [doi: 10.3390/cancers2041980] [Medline: 24281212]
3. Haenssle H, Fink C, Schneiderbauer R, Toberer F, Buhl T, Blum A, Reader Study Level-I and Level-II Groups. Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Ann Oncol 2018 Aug 01;29(8):1836-1842. [doi: 10.1093/annonc/mdy166] [Medline: 29846502]
4. Argenziano G, Soyer HP. Dermoscopy of pigmented skin lesions: A valuable tool for early diagnosis of melanoma. Lancet Oncol 2001 Jul;2(7):443-449. [Medline: 11905739]
5. Kittler H, Pehamberger H, Wolff K, Binder M. Diagnostic accuracy of dermoscopy. Lancet Oncol 2002 Mar;3(3):159-165. [Medline: 11902502]
6. Ali ARA, Deserno TM. A systematic review of automated melanoma detection in dermatoscopic images and its ground truth data. Proc SPIE Int Soc Opt Eng 2012 Feb 28;8318:1-6. [doi: 10.1117/12.912389]
7. Fabbrocini G, De Vita V, Pastore F, D'Arco V, Mazzella C, Annunziata MC, et al. Teledermatology: From prevention to diagnosis of nonmelanoma and melanoma skin cancer. Int J Telemed Appl 2011 Sep 01;2011(17):125762 [FREE Full text] [doi: 10.1155/2011/125762] [Medline: 21776252]
8. Foraker R, Kite B, Kelley MM, Lai AM, Roth C, Lopetegui MA, et al. EHR-based visualization tool: Adoption rates, satisfaction, and patient outcomes. EGEMS (Wash DC) 2015;3(2):1159 [FREE Full text] [doi: 10.13063/2327-9214.1159] [Medline: 26290891]
9. Fabbrocini G, Betta G, Di Leo G, Liguori C, Paolillo A, Pietrosanto A, et al. Epiluminescence image processing for melanocytic skin lesion diagnosis based on 7-point check-list: A preliminary discussion on three parameters. Open Dermatol J 2010 Jan 01;4(1):110-115. [doi: 10.2174/1874372201004010110]
10. Hart PE, Stork DG, Duda RO. Pattern Classification. 2nd edition. Hoboken, NJ: John Wiley & Sons; 2000.
11. Oliveira RB, Papa JP, Pereira AS, Tavares JMRS. Computational methods for pigmented skin lesion classification in images: Review and future trends. Neural Comput Appl 2016 Jul 15;29(3):613-636. [doi: 10.1007/s00521-016-2482-6]
12. Gutman D, Codella NCF, Celebi E, Helba B, Marchetti M, Mishra N, et al. arXiv. 2016 May 04. Skin lesion analysis toward melanoma detection: A challenge at the International Symposium on Biomedical Imaging (ISBI) 2016, hosted by the International Skin Imaging Collaboration (ISIC) URL: https://arxiv.org/pdf/1605.01397 [accessed 2018-10-06] [WebCite Cache ID 72yoIScsz]
13. Marchetti MA, Codella NCF, Dusza SW, Gutman DA, Helba B, Kalloo A, International Skin Imaging Collaboration. Results of the 2016 International Skin Imaging Collaboration International Symposium on Biomedical Imaging challenge: Comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images. J Am Acad Dermatol 2018 Dec;78(2):270-277.e1. [doi: 10.1016/j.jaad.2017.08.016] [Medline: 28969863]
14. Demyanov S, Chakravorty R, Abedini M, Halpern A, Garnavi R. Classification of dermoscopy patterns using deep convolutional neural networks. In: Proceedings of the 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI).: IEEE; 2016 Presented at: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI); April 13-16, 2016; Prague, Czech Republic p. 2-12.
15. Carcagni P, Cuna A, Distante C. arXiv. 2018 Jul 26. A dense CNN approach for skin lesion classification URL: http://arxiv.org/pdf/1807.06416 [accessed 2018-10-06] [WebCite Cache ID 72ypBqU6n]
16. Dorj U, Lee K, Choi J, Lee M. The skin cancer classification using deep convolutional neural network. Multimed Tools Appl 2018 Feb 22;77(8):9909-9924. [doi: 10.1007/s11042-018-5714-1]
17. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015 May 28;521(7553):436-444. [doi: 10.1038/nature14539] [Medline: 26017442]

18. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 2012 Jan:1-23. [doi: 10.1145/3065386]
19. Simonyan K, Zisserman A. arXiv. 2014 Apr 10. Very deep convolutional networks for large-scale image recognition URL: https://arxiv.org/pdf/1409.1556 [accessed 2018-10-06] [WebCite Cache ID 72yq41Xea]
20. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. arXiv. 2014 Sep 17. Going deeper with convolutions URL: https://arxiv.org/pdf/1409.4842 [accessed 2018-10-06] [WebCite Cache ID 72yqGoKqD]
21. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016 Presented at: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); June 27-30, 2016; Las Vegas, NV.
22. Menegola A, Fornaciali M, Pires R, Vasques Bittencourt F, Avila Fontes de Avila S, Valle E. Knowledge transfer for melanoma screening with deep learning. In: Proceedings of the 14th IEEE International Symposium on Biomedical Imaging, ISBI 2017. 2017 Presented at: 14th IEEE International Symposium on Biomedical Imaging, ISBI 2017; April 18-21, 2017; Melbourne, Australia p. 297-300. [doi: 10.1109/ISBI.2017.7950523]
23. Pomponiu V, Nejati H, Cheung NM. Deepmole: Deep neural networks for skin mole lesion classification. In: Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP). 2016 Presented at: 2016 IEEE International Conference on Image Processing (ICIP); September 25-28, 2016; Phoenix, AZ.
24. Codella N, Cai J, Abedini M, Garnavi R, Halpern A, Smith JR. Deep learning, sparse coding, and SVM for melanoma recognition in dermoscopy images. In: Proceedings of the 6th International Workshop on Machine Learning in Medical Imaging. 2015 Presented at: 6th International Workshop on Machine Learning in Medical Imaging; October 5-9, 2015; Munich, Germany p. 118-126. [doi: 10.1007/978-3-319-24888-2_15]
25. Kawahara J, BenTaieb A, Hamarneh G. Deep features to classify skin lesions. In: Proceedings of the 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI). 2016 Presented at: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI); April 13-16, 2016; Prague, Czech Republic. [doi: 10.1109/ISBI.2016.7493528]
26. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017 Dec 02;542(7639):115-118. [doi: 10.1038/nature21056] [Medline: 28117445]
27. Han SS, Kim MS, Lim W, Park GH, Park I, Chang SE. Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. J Invest Dermatol 2018 Jul;138(7):1529-1538. [doi: 10.1016/j.jid.2018.01.028] [Medline: 29428356]
28. Bi L, Kim J, Ahn E, Feng D. arXiv. 2017. Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks URL: http://arxiv.org/pdf/1703.04197 [accessed 2018-10-07] [WebCite Cache ID 72zjKXpbC]
29. Kawahara J, Hamarneh G. Multi-resolution-tract CNN with hybrid pretrained and skin-lesion trained layers. In: Proceedings of the 7th International Conference on Machine Learning in Medical Imaging (MLMI 2016). 2016 Presented at: 7th International Conference on Machine Learning in Medical Imaging (MLMI 2016); October 17, 2016; Athens, Greece.
30. Sun X, Yang J, Sun M, Wang K. A benchmark for automatic visual classification of clinical skin disease images. In: Proceedings of the European Conference on Computer Vision. 2016 Presented at: European Conference on Computer Vision; October 8-16, 2016; Amsterdam, The Netherlands.
31. Lopez AR, Giro-i-Nieto X, Burdick J, Marques O. Skin lesion classification from dermoscopic images using deep learning techniques. In: Proceedings of the 13th IASTED International Conference on Biomedical Engineering. 2017 Presented at: 13th IASTED International Conference on Biomedical Engineering; February 20-21, 2017; Innsbruck, Austria. [doi: 10.2316/P.2017.852-053]
32. Nasr-Esfahani E, Samavi S, Karimi N, Soroushmehr SMR, Jafari MH, Ward K, et al. Melanoma detection by analysis of clinical images using convolutional neural network. In: Proceedings of the 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2016 Presented at: 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); August 16-20, 2016; Orlando, FL.
33. Navarrete-Dechent C, Dusza SW, Liopyris K, Marghoob AA, Halpern AC, Marchetti MA. Automated dermatological diagnosis: Hype or reality? J Invest Dermatol 2018 Oct;138(10):2277-2279. [doi: 10.1016/j.jid.2018.04.040] [Medline: 29864435]

Abbreviations

AUC: area under the curve
CNN: convolutional neural network
FN: false negative
FP: false positive
ISBI: International Symposium on Biomedical Imaging
ISIC: International Skin Imaging Collaboration
ROC: receiver operating characteristic
TN: true negative
TP: true positive

Edited by G Eysenbach; submitted 14.08.18; peer-reviewed by V De Vita, G Fabbrocini; comments to author 05.09.18; revised version received 05.09.18; accepted 08.09.18; published 17.10.18.

Please cite as:
Brinker TJ, Hekler A, Utikal JS, Grabe N, Schadendorf D, Klode J, Berking C, Steeb T, Enk AH, von Kalle C
Skin Cancer Classification Using Convolutional Neural Networks: Systematic Review
J Med Internet Res 2018;20(10):e11936
URL: http://www.jmir.org/2018/10/e11936/
doi:10.2196/11936
PMID:30333097

Original Paper

Designing a Patient Portal for Patient-Centered Care: Cross-Sectional Survey

Steve Alfons Van den Bulck[1], MD; Rosella Hermens[1,2], PhD; Karin Slegers[3], PhD; Bert Vandenberghe[3], MSc; Geert Goderis[1], MD, PhD; Patrik Vankrunkelsven[1], MD, PhD

[1] Academic Center for General Practice, Department of Public Health and Primary Care, KU Leuven, Leuven, Belgium

[2] Scientific Institute for Quality in Healthcare, Radboud Institute for Health Sciences, Radboud University Medical Center, Radboud University Nijmegen, Nijmegen, Netherlands

[3] Meaningful Interactions Lab, Imec, KU Leuven, Leuven, Belgium

Corresponding Author:
Steve Alfons Van den Bulck, MD
Academic Center for General Practice
Department of Public Health and Primary Care
KU Leuven
Kapucijnenvoer 33 Blok J
Leuven, 3000
Belgium
Phone: 32 16 37 90 14
Fax: 32 16 33 74 80
Email: steve.vandenbulck@kuleuven.be

Abstract

Background: In recent literature, patient portals are considered as important tools for the delivery of patient-centered care. To date, it is not clear how patients would conceptualize a patient portal and which health information needs they have when doing so.

Objective: This study aimed (1) to investigate health information needs, expectations, and attitudes toward a patient portal and (2) to assess whether determinants, such as patient characteristics, health literacy, and empowerment status, can predict two different variables, namely the importance people attribute to obtaining health information when using a patient portal and the expectations concerning personal health care when using a patient portal.

Methods: We conducted a cross-sectional survey of the Flemish population on what patients prefer to know about their digital health data and their expectations and attitudes toward using a patient portal to access their electronic health record. People were invited to participate in the survey through newsletters, social media, and magazines. We used a questionnaire including demographics, health characteristics, health literacy, patient empowerment, and patient portal characteristics.

Results: We received 433 completed surveys. The health information needs included features such as being notified when one's health changes (371/396, 93.7%), being notified when physical parameters increase to dangerous levels (370/395, 93.7%), observing connections between one's symptoms or diseases or biological parameters (339/398, 85.2%), viewing the evolution of one's health in function of time (333/394, 84.5%), and viewing information about the expected effect of treatment (349/395, 88.4%). Almost 90% (369/412) of respondents were interested in using a patient portal. Determinants of patients' attachment for obtaining health information on a patient portal were (1) age between 45 and 54 years (P=.05); (2) neutral (P=.03) or interested attitude (P=.008) toward shared decision making; and (3) commitment to question physicians' decisions (P=.03, R^2=0.122). Determinants of patients' expectations on improved health care by accessing a patient portal were (1) lower education level (P=.04); (2) neutral (P=.03) or interested attitude (P=.008) toward shared decision making; and (3) problems in understanding health information (P=.04; R^2=0.106).

Conclusions: The interest in using a patient portal is considerable in Flanders. People would like to receive alerts or some form of communication from a patient portal in case they need to act to manage their health. Determinants such as education, attached importance to shared decision making, difficulties in finding relevant health information, and the attached importance in questioning the decisions of physicians need to be considered in the design of a patient portal.

(J Med Internet Res 2018;20(10):e269) doi:10.2196/jmir.9497

KEYWORDS

cross-sectional studies; consumer health informatics; electronic health record; patient portal; personal health records

Introduction

The expression “no decision about me, without me,” as formulated by the British National Health Service, emphasizes the importance of patient-centered care and shared decision making [1]. Patient-centered care, which is an important feature of high-quality health care, is defined as “providing care that is respectful of and responsive to individual patient preferences, needs, and values, and ensuring that patient values guide all clinical decisions” [2]. The Institute of Medicine considers patient-centered care as one of the 6 objectives to be achieved to improve the quality of US health care [2]. It can be promoted with the help of patient-centered health information technologies [3], such as a patient portal.

A patient portal is known to improve the quality of and access to health care [4]. In addition, a patient portal exerts a positive influence on health care utilization [5]. A patient portal can be defined as “An electronic application through which individuals can access, manage and share their health information, and that of others for whom they are authorized, in a private, secure, and confidential environment” [6,7]. More specifically, a patient portal can be described as an application that is linked (tethered) to the electronic health record of the doctor [8,9]. The functions of a patient portal often include a medication list, test results, a list of allergies, a problem list, social history, major illness, lifestyle, family history, and links to personalized information [7,9].

Although patient portals are considered important tools for the development of patient-centered care, the current use is not optimal, and portals are still less patient-centered than they could be [10-12]. Known barriers to the use of portals for patients and providers include security and privacy issues, the negative impact on the workflow and limited user-friendliness [4,13]. Benefits associated with the use of a patient portal are increased convenience and satisfaction for patients [14,15].

Patient safety could be improved by identifying errors in medication lists [16,17]. Moreover, providing Web access could lower the threshold for the patient-clinician contact [13]. In addition, the quality of care can be improved by the sustained use of some features of a patient portal. For example, Web-based prescription refills and secure messaging have an impact on some physiological measures important for the management of type 2 diabetes [18].

Patients’ interest and ability to use a patient portal are influenced by age, health literacy, and level of education. Patients are more likely to adopt a patient portal if its features align with their information needs and with the functionalities they require [19].

To date, however, it remains unclear what patients’ information needs and functionality requirements exactly are. In addition, participants in this study had no prior experience with patient portals as they do not exist in Belgium. The survey intro was the first exposure people had to the concept of a patient portal. Therefore, this study aimed to investigate patients’ information needs with regard to the concept of a patient portal and its presumed use to access their health data. Furthermore, this study aims to assess patients’ expectations and attitudes regarding a patient portal and its use to access electronic health records in primary care. Finally, we investigate whether patients’ information needs, expectations, and attitudes regarding a patient portal correlate with patient-related determinants such as personal characteristics, health literacy, and empowerment status.

Methods

Study Design

We conducted a cross-sectional survey on the digital health information needs of patients and their expectations and attitudes toward using a patient portal to access their electronic health record in primary care among inhabitants of the Dutch-speaking part of Belgium (Flanders).

Study Population and Sample Size

The study population included inhabitants of the Flemish part of Belgium. For precise estimation of our results, we calculated the sample size using the free Web-based software Raosoft [20]. For a population size of 6,471,996, with a 5% margin of error and a 50% response distribution, we needed 385 completed surveys.

Design and Domains of the Survey

Our open and voluntary survey was a convenience sample and consisted of 2 parts (Multimedia Appendix 1 for the English version and Multimedia Appendix 2 for the Dutch Version).

The first part was based on the Health Information Technology Evaluation Collaborative (HITEC) Consumer Health Information Technology Survey as designed by Patel et al [21]. The second part was based on findings from previous qualitative research. The first page of the survey contained the informed consent and a mandatory checkbox for participants to acknowledge that they have read and agree with the informed consent. The final page of the survey contained a mandatory checkbox for participants to acknowledge they wanted to send the information to the researchers. If not, the data were not stored. Multimedia Appendix 3 provides a detailed description of the survey methodology according to the Checklist for Reporting Results of Internet E-Surveys [22].

Part 1 of the Survey

The HITEC Consumer Survey, developed in the United States, aimed to characterize consumer attitudes toward personal health records and included questions on potential personal health record use, preferences, and attitudes toward personal health records, the use of computers and the internet, experience with health care, health characteristics, and demographics [21].

One researcher translated the HITEC Consumer Survey from English to Dutch, and 2 researchers with adequate English proficiency translated the Dutch survey independently from each other back to English to reduce language bias to a minimum [23,24]. Small differences in the translation were solved by consensus. In the Dutch translation, a patient portal was defined as "an electronic online tool to view and manage patients' health care information stored in the electronic health record of their general physician" [8]. Of note, 2 questions were not included in the Dutch translation of the questionnaire—one about payment options and one about private insurance companies—because these were not applicable to the Belgian situation. A few other questions were slightly changed because of the different demographic situation in Belgium compared with that in the United States. For example, questions inquiring about ethnicity were reformulated to match the common Dutch terminology. Instead, of asking "what is your race" or "are you of Latino or Hispanic origin or descent," we asked, "What is your mother's, father's, and your country of birth?"

Part 2 of the Survey

The questions in this part of the survey were based on the findings of a previous qualitative study (Multimedia Appendix 4). This study aimed to understand what motivated people to search for health information. For this purpose, cultural probes and the Lillidots method were used [25]. Cultural probes are objects, such as a diary or camera, provided to participants to self-report data; cultural probes perform similar to astronomic probes because they are also left behind and return with fragmentary data [26,27]. In the study mentioned above, cultural probes were used to sensitize participants about digital health data to make it easier for them to express their feelings and experiences regarding health information. Participants were then interviewed to learn what health topics they wanted to know more about and what their questions and expectations were concerning these topics. For this purpose, an ideation method, called the Lillidots method [25], was combined with a strategy to anthropomorphize the technology at hand (digital health data in this case). This method, detailed elsewhere [28], resulted in insights about the type of health data people are interested in, questions they have in this respect, and their motivations to ask these questions. The second part of the survey was based on these motivations and included questions such as "In your relationship with your doctor, how important is it for you to be able to question their decision" and "Please indicate how important each of the following aspects would be to you, if you had access to all the data in your patient portal."

In a nutshell, the second part of the survey was developed based on the themes that emerged from the previous qualitative study: (1) health awareness; (2) coping; (3) effective care; (4) empowerment; (5) good health; (6) patient rights; and (7) getting recognition. These themes grouped several items, such as people want to understand whether and when care is necessary and compare possible treatments (effective care) or what reimbursements they are entitled to (patient rights). To validate these findings with a larger population a questionnaire was constructed based on these items (part 2 of the survey). Combining part 1 and 2 allowed us to use a quantitative research instrument with our bottom-up qualitative research method.

Pilot

A pilot was performed in a small sample of 6 Dutch-speaking Belgians to establish whether the survey was feasible, particularly regarding the ease of use and time of completion. All participants appraised the survey as user-friendly and completed it in an average time of 20 minutes, which was considered acceptable. We did not include any pilot data in the large sample for analysis.

Data Collection

The data were collected between March 25 and September 1, 2016. Inhabitants of the Flemish part of Belgium were invited to participate through several channels—the member magazine of a health insurance with a print run expanding up to 1,200,000 copies, only this channel gave readers an opportunity to choose between Web-based reply and paper; the Web-based newsletter of the same health insurance; the newsletter of a website called "Health and Science," which is an independent and evidence-based website to inform patients about health-related topics (12,000 subscribers); and the website of a well-known Flemish weekly magazine (knack.be), which receives 54,769 unique visitors and 141,304 (2.58 per visitor) page views per day [29] and their social media channels (Facebook and Twitter). In addition, some students of the Faculty of Social Sciences used their social networks to recruit participants. Two reminders were sent by the health insurance and the "Health and Science" newsletter to obtain the greatest possible response. People without access to a computer or the internet could phone to request a paper copy of the survey. This ensured involving as many people as possible and having a representative sample of the population to minimize selection bias. The paper surveys we received were manually entered into the database containing the results from the Web-based surveys.

Analyses

Descriptive Statistics (Frequencies)

In this study, descriptive statistics (frequencies) were used to assess the following:

- Demographics (age, gender, mother tongue, employment status, education, income, internet use, and residence),
- Health characteristics (answers to questions on self-rated health status, the presence of a chronic disease, >3 annual visits to the primary care physician, and prescribed medication use),
- Health literacy (answers to questions on difficulties with finding relevant and reliable health information),
- Patient empowerment (answers to questions on the quality of health care received in the last 5 years and questions about the importance of shared decision making and being able to question physicians' decisions),
- The potential use of a patient portal, attitude and expectations when using a patient portal (answers to questions on one's interest in using a patient portal, types of information expected to be found in a patient portal, Web-based activities concerning health, frequency of using a patient portal, perceived usefulness of a patient portal, difficulty with using a patient portal, granting access to

one's own patient portal, and importance of certain features of a patient portal).

Determinant Analyses

To investigate whether determinants such as patient characteristics, health literacy, and patient empowerment predict the importance people attach to obtaining health information when accessing a patient portal and the expectations and attitudes of people toward the use of a patient portal, we used linear and logistic regression models. First, bivariate regression models were used to investigate associations between all variables. Only predictor variables that were significantly associated ($P \leq .10$) with the dependent variable in bivariate analyses were used for the multivariate analyses.

Bivariate and multivariate linear regression models were used to predict 2 different dependent variables. For one of these "the importance people attach to obtaining health information when using a patient portal," we used the mean sum score of 5-point Likert scale answers (ranging from 1 "very important" to 5 "very unimportant") to 12 questions asking about the importance of doing certain things with the help of a patient portal (Cronbach alpha=.897). We considered the mean sum score as an interval variable and used it as a dependent variable in a linear regression model. [30]

For the second dependent variable "expectations concerning one's personal health care when using a patient portal," we used the mean sum score of answers to 9 questions asking for the level of expected improvement on different aspects of health care when using a patient portal (ranging from 1 "will greatly improve" to 5 "will greatly worsen"; Cronbach alpha=.871). We used the mean sum score to compensate for missing data in ≥1 of the questions. Preliminary analyses were performed to ensure there was no violation of the assumption of normality, linearity, and multicollinearity. If a correlation (>.6) between independent variables was detected, only the variable with the greatest influence on the $R^2$of the model was included in the multivariate analyses.

We used bivariate and multivariate logistic regression models to predict the attitude toward using a patient portal. We used answers to the question asking for the interest in using a patient portal as the dependent variable. This ordinal variable with 5 categories was recoded into a dichotomous variable (interested and not interested or neutral)

As predictor variables, we used answers to the questions asking for patient characteristics (age, employment status, family income, health status, education, and gender), patient empowerment (shared decision making, questioning the decisions of physicians, satisfaction with health care received the last 5 years), and health literacy (finding relevant information about health, evaluating the reliability of health information, and problems in understanding health information; Multimedia Appendix 5).

Except for the internal consistency (Cronbach alpha) of our composite dependent variables, we did not include any psychometric measures for our predictor variables because of the gap in the availability of sound psychometric measures for evaluating patient-facing eHealth technologies [31]. All statistical analyses were performed with IBM SPSS Statistics version 24 for Windows (IBM Corp, Armonk, NY, USA).

Missing Data

To calculate the frequencies, missing data were excluded, and percentages were based on the number of nonmissing values. For the regression models, we used independent variables that had a low percentage of missing data (<15%). If values of any of the independent variables included missing data, the entire case was excluded for the analysis.

Ethical Approval

This study was approved by the Social and Societal Ethics Committee of the Faculty of Social Sciences of KU Leuven on July 14, 2015 (first part of the survey) and on September 24, 2015 (second part of the survey) with the grant number G-2015 07 272.

Results

Demographics and Health Characteristics

The survey was completed by 433 people. While 10 surveys were submitted on paper, 423 were completed through the Web (Table 1). The completion rate of the Web-based survey was 91% (Multimedia Appendix 3). Differences between the 10 surveys submitted on paper and those completed on the Web were mainly age of respondents, age-related properties (eg, employment status), and their internet use. The mean age in the group of the paper surveys (n=10) was 68.6 years (SD 10.146), while it was 53 years (SD 16.497) in the Web-based group (n=423). All respondents of the paper surveys group were not working anymore (9/10 retired and 1/10 disabled), and 40% (4/10) did rarely or never used the internet, while only 0.2% (1/423) of the Web-based group rarely or never used the internet. The mean age of all the respondents (n=433) was 53.28 (SD 16.451) years.

Almost 92% (397/432) of participants reported that they were in excellent or good health, although 49.4% (213/431) reported they had a chronic disease, and 69.4% (300/432) were taking prescribed medication (Table 1).

Health Literacy and Patient Empowerment

Finding relevant health information was considered difficult by 26.2% (106/404) of respondents, and assessing the reliability of the health care information found was considered difficult by 48.3% (195/404). More than 93% (377/404) of participants reported that they found shared decision making important, and 89.9% (363/404) of respondents thought it was important to be able to question the decisions made by physicians (Table 2).

Table 1. Demographics and health characteristics of the participants (N=433).

Characteristics	Participants, n (%)[a]
Male	187 (47.1)
Dutch-speaking	394 (99.0)
Age (years)	
18-34	72 (19.1)
35-44	39 (10.4)
45-54	53 (14.1)
55-64	95 (25.3)
>65	117 (31.1)
Employment status	
Employed	164 (41.3)
Student	21 (5.3)
Unemployed	51 (12.8)
Retired	161 (40.6)
Education	
Secondary school	130 (32.7)
Bachelor degree	159 (40.1)
Master degree or higher	108 (27.2)
Family income	
€ <30,000	114 (28.9)
€ 30,000-60,000	100 (25.3)
€ >60,000	58 (14.7)
Does not know or prefers not to disclose	123 (31.1)
Internet use	
Internet use ≥1 time per day	423 (97.9)
Searched information on health or disease on the Web	406 (94.4)
Residence description	
Rural	202 (50.6)
Urban	197 (49.4)
Self-rated health status	
Excellent or very good	121 (28.0)
Good or fair	276 (63.9)
Poor	35 (8.1)
Chronic medical condition	213 (49.4)
Visited primary caregiver >3 times in a year	214 (50.1)
Taking prescribed medication	300 (69.4)

[a] Numbers may not sum to totals because of missing data. Percentages were calculated considering the missing data.

Table 2. Health literacy and patient empowerment (N=433).

Health literacy and patient empowerment	Participants, n (%)[a]
Satisfaction with the quality of health care received in the last 5 years	
Satisfied	366 (84.5)
Neutral	39 (9.0)
Dissatisfied	28 (6.5)
Finding relevant health information	
Difficult	106 (26.2)
Not easy or not difficult	149 (36.9)
Easy	149 (36.9)
Assessing the reliability of Web-based health care information	
Difficult	195 (48.3)
Not easy or not difficult	116 (28.7)
Easy	93 (23.0)
Importance of shared decision making	
Important	377 (93.3)
Neutral	18 (4.5)
Not important	9 (2.2)
Importance of being able to questions physicians' decisions	
Important	363 (89.9)
Neutral	37 (9.1)
Not important	4 (1.0)

[a] Numbers may not sum to totals because of missing data. Percentages were calculated considering the missing data.

Patient Portal's Potential Impact and Features

Most respondents were interested in the potential use of a patient portal (369/412, 89.6%; Table 3). The information that most people wanted to see in their patient portal were test results (381/410, 92.9%), current medication (345/410, 84.1%), immunization records (338/410, 82.4%), and their past medical visits, procedures, and surgeries (338/410, 82.4%). This corresponds with the Web-based health-related activities people are most interested in, namely viewing their medical records, test results, medication list (384/405, 94.8%); requesting appointments, referrals, and prescription refills (376/403, 93.3%); and signing up for reminders for preventive medicine (360/399, 90.2%).

Respondents were less interested in seeing their lifestyle choices (138/410, 33.7%) and information from devices to help monitor their health (199/410, 48.5%; Table 3).

The perceived impact of patient portal use varied. Only 22.3% (90/404) of respondents believed that the patient portal use would improve the security and privacy of their medical data, and 47.4% (192/405) of participants thought that using a patient portal would reduce the overall cost of their health care.

The majority (391/414, 94.4%) would give their primary care doctor permission to view information in their patient portal. The potential features of a patient portal that were considered important by respondents were being notified when certain physical parameters evolve toward dangerous levels (370/395, 93.7%), being notified when their health changes (371/396, 93.7%), being able to view the expected impact of treatment on personal health (349/395, 88.4%), being able to see connections between symptoms, disease(s), biological parameters, etc (339/398, 85.2%), and being able to view the evolution of their health in function of time (333/394, 84.5%). Consumers were less interested in comparing their personal health data with anonymous data from other patients (146/396, 36.9%) and with anonymous data from the Flemish population (146/394, 37.1%).

Determinant Analysis of the Importance People Attribute to Obtaining Health Information When Using a Patient Portal

Bivariate linear regression showed that the importance people attributed to obtaining health information when using a patient portal to access health data (dependent) was significantly associated (cutoff $P \leq .10$) with age, employment status, self-rated health status, the interest in shared decision making, the importance of being able to question the decisions of physicians, the difficulty in finding relevant health information, the difficulty in assessing the reliability of health information, and the difficulty in understanding health information. Due to the collinearity between age and employment status, only age was used for multivariate analysis.

Table 3. Patient portal use characteristics (N=433).

Patient portal characteristics	Participants, n (%)[a]
Interested in using a patient portal	369 (89.6)
Types of information people would prefer to have in their patient portal	
My allergies	263 (64.1)
Test results (eg, X-rays, blood tests, etc)	381 (92.9)
Immunization records (list of vaccines received)	338 (82.4)
Medication I have taken or am currently taking	345 (84.1)
List of doctors or health care providers I have seen	279 (68.0)
Family history of health problems	236 (57.6)
Medical problems	322 (78.5)
Medical visits or surgeries or medical procedures I have had	338 (82.4)
Lifestyle choices (eg, smoking history and exercise)	138 (33.7)
Information from devices that help me monitor my health	199 (48.5)
Activities I am doing or would like to do on the internet	
View medical records, test results, and medication list	384 (94.8)
Add notes to my medical record	312 (78.6)
Request appointments, referrals, prescription refills	376 (93.3)
Communicate with my doctor and receive reports by mail	358 (88.8)
Fill out paperwork before or after a physician visit	331 (83.0)
Sign up for reminders for preventive medicine (eg, flu shot)	360 (90.2)
Learn about opportunities to participate in medical research	327 (82.6)
Access my child’s or parents’ medical record if I am primary caretaker	315 (81.2)
Communicate with other people with similar health problems	197 (50.8)
Receive educational materials related to my health	304 (79.4)
Record my representative to manage my health care when I am not able	337 (83.6)
Expected frequency of patient portal use	
At least 1 time per week	61 (14.8)
1 time per month	159 (38.6)
Every 3-6 months	159 (38.6)
Rarely or never	33 (8.0)
How many people think the use of a patient portal will improve the following[b]	
Security and privacy of my medical information	90 (22.3)
Communication between my doctors and myself	321 (78.7)
My understanding of my own health	300 (73.7)
My sense of control over my own health care	311 (76.4)
My worries about my own health care	236 (58.1)
The safety of my care (freedom from errors)	278 (68.6)
My satisfaction with my health care	275 (68.1)
The overall quality of my health care	292 (72.1)
The overall costs of my health care	192 (47.4)
Difficulty to use a patient portal to view and manage your health information and care	
Difficult	22 (5.3)

Patient portal characteristics	Participants, n (%)[a]
Easy	292 (70.9)
Neutral	98 (23.8)
Who would you give permission to view information in your patient portal	
Designated family members or friends	226 (54.6)
My primary care doctor	391 (94.4)
Other doctors or health care providers who care for me	332 (80.2)
My health insurance	26 (6.3)
My employer	1 (0.2)
The government	8 (1.9)
No-one	18 (4.3)
Importance of certain features when using a patient portal[c]	
Compare recent personal health data with health data from the past	311 (78.1)
Compare personal health data with medical standards	291 (73.5)
Compare personal health data with anonymous data from other patients	146 (36.9)
Compare personal health data with anonymous data from the Flemish population	146 (37.1)
See connections between your symptoms, your disease(s), your biological parameters	339 (85.2)
See connections between your health and the presence of environmental factors	284 (71.7)
View the evolution of your health in function of time	333 (84.5)
View information about the expected effect of treatment on your personal health	349 (88.4)
View information on the expected impact of your lifestyle on your personal health	309 (79.0)
Provide your data anonymously so that regional problems can be detected	250 (63.1)
Be notified when certain physical parameters evolve toward dangerous levels	370 (93.7)
Be notified when your health changes	371 (93.7)

[a] Numbers may not sum to totals due to missing data. Percentages were calculated considering the missing data.

[b] Used as a composite dependent variable: “expectation concerning one’s personal health care when using a patient portal.”

[c] Used as a composite dependent variable: “importance attached to obtaining health information when using a patient portal to access health data.”

In the multivariate linear regression, age, shared decision making, and the importance of being able to question the decisions of physicians were significant (cutoff $P \leq .05$). People attributed greater importance to receiving health information if they were aged 45-54 years compared with those who were aged 18-34 years, if they had an interested or neutral attitude toward shared decision making compared with having a negative attitude toward shared decision making and if they found it important to be able to question the decisions made by physicians compared with finding this unimportant ($R^2=0.122$; Table 4).

Determinant Analysis of Expectations Concerning Personal Health Care When Using a Patient Portal

Bivariate linear regression showed that expectations concerning personal health care when using a patient portal (dependent) were significantly associated (cutoff $P \leq .10$) with age, employment status, education, the interest in shared decision making, the importance of being able to question the decisions of physicians, the difficulty in finding relevant health information, the difficulty in assessing the reliability of health information, and the difficulty in understanding health information.

Due to the collinearity between age and employment status, only the employment status was used for multivariate analysis. Multivariate linear regression showed that expectations concerning personal health care when using a patient portal (dependent) was significantly associated (cutoff $P \leq .05$) with education, shared decision making, the difficulty in finding relevant health information and problems in understanding health information (Table 4). People expected an improvement in their individual health care when using a patient portal if they had lower levels of education (high school degree or lower) compared with highly educated participants (master degree or higher), an interested or neutral attitude toward shared decision making compared with a negative attitude toward shared decision making and if they sometimes had problems to understand health information compared with rarely having problems in understanding health information. People expected an impairment in their health care when using a patient portal if they found it easy or had a neutral attitude toward finding relevant health information compared with thinking it is difficult to find this information (R^2=0.106; Table 4).

Determinant Analysis of the Interest in Using a Patient Portal

Owing to the low variance between the independent and dependent variables, the determinant analysis showed results with very high uncertainty and was not conclusive.

Table 4. Multivariate regression models.

Multivariate regression models	*P* value	Beta (95% CI)
Importance of obtaining health information		
Shared decision making		
Important	.03	–.558 (–0.969 to –0.147)
Neutral	.008	–.500 (–0.952 to –0.048)
Unimportant (constant)	—[a]	—
Questioning decisions of physicians		
Important	.03	–.642 (–1.222 to –0.061)
Neutral	.12	–.467 (–1.052 to 0.119)
Unimportant (constant)	—	—
Age[b] in years		
18-34 (constant)	—	—
35-44	.069	–.186 (–0.386 to 0.014)
45-54	.047	–.183 (–0.363 to –0.002)
55-64	.10	–.126 (–0.278 to 0.026)
>65	.099	–.124 (–0.271 to 0.023)
Expectations concerning personal health care		
Education		
High school or lower	.04	–.155 (–0.303 to –0.007)
Bachelor degree	.84	–.015 (–0.153 to 0.124)
Master degree or higher (constant)	—	—
Shared decision making		
Important	.008	–.566 (–0.983 to –0.148)
Neutral	.03	–.521 (–0.989 to –0.052)
Unimportant (constant)	—	—
Finding relevant health information		
Difficult (constant)	—	—
Not difficult or not easy	.017	.180 (0.032 to 0.328)
Easy	.022	.197 (0.028-0.366)
Problems in understanding health information		
Frequently	.99	.002 (–0.234 to 0.237)
Sometimes	.037	–.141 (–0.272 to –0.009)
Rarely (constant)	—	—

[a] Not applicable.

[b] Owing to collinearity with employment status, only age was used for the multivariate regression.

Discussion

Principal Findings

This study used a cross-sectional survey design to investigate health information needs, expectations, and interest of people accessing a patient portal to view their health data. The health information needs in this context are mainly features such as being notified when one’s health changes and being notified when physical parameters increase to dangerous levels. In addition, people are interested to see connections among their symptoms, disease, and biological parameters. Moreover, they prefer to have an overview of how their health evolves and to have access to information about the expected effect of treatment on their health. These findings are in line with the patients’ desire for direct communication through a patient portal that was described in previous work [32].

The importance that people attribute to obtaining health information when using a patient portal is partially predicted by age and the level of patient empowerment, namely the importance of shared decision making and questioning physicians’ decisions. Expectations concerning the impact on personal health care when using a patient portal are influenced by the level of education, interest in shared decision making, and the difficulty people experience in finding and understanding the relevant health care information. These findings confirm that screening for eHealth literacy and providing training in the use of a patient portal could help in improving the experience and expectations people have when using a patient portal [33,34].

People expect that the use of a patient portal can improve the communication between their physician and themselves. However, only a few people think that the use of a patient portal will improve the security and privacy of their medical information. In this study, the expressed interest in using a patient portal was high, with almost 90% of respondents interested in the use of a patient portal, although there was no functional patient portal available in Belgium at the time the survey was submitted. The fact that people in Flanders express the need to be notified when their health changes highlights the interest of people in receiving some form of communication or alert through their patient portal as to when they need to act to manage their health care. This corresponds with previous research that states that receiving intelligent alerts is an important feature in the conceptual design of an integrated shared decision personal health record [35]. A patient-directed clinical decision system that is integrated into a patient portal could be useful for this purpose as the ongoing research hypothesizes [36].

Although earlier research shows that incorporating patients’ lifestyle is important for patient portals to become more user-centric [37], only a third of our respondents were interested to have data about lifestyle choices in their patient portal. A potential explanation might be that people when conceptualizing a patient portal consider themselves to be aware of their lifestyle and therefore do not feel the need to find this information in a patient portal. Patients in Flanders, who have no experience with a patient portal, might think about a patient portal as a unidirectional channel where they can find and consult their health information, where people who have actually used a patient portal, value it as a bidirectional tool [11].

Another important finding is the fact that almost 9 of 10 respondents consider it important to be able to question the decisions made by physicians. This finding, together with the importance of shared decision making (93% consider this important), emphasizes the fact that people in Flanders are critical health care users who greatly appreciate patient empowerment and endorses the results from a former study where Belgian inhabitants attained fairly high empowerment scores [38].

Previous research showed that people think the use of a patient portal can improve the communication between their physician and themselves [39], the understanding of their own health [14], the sense of control over their health care [40], and the overall quality of their health care [4]. In literature, there is some skepticism about the influence of a patient portal on the security and privacy of health data and the total costs of health care [13,41,42]. People are mostly interested in test results, medication schemes, immunization records, and a history of medical visits and procedures [21,43]. These findings are consistent with ours, where about three-quarters of respondents believed that a patient portal could improve the doctor-patient communication, the understanding and sense of control of their health, and the quality of care.

Only a small percentage of respondents (22.3%) think the use of a patient portal can improve the privacy and security of medical information and almost half of them believe a patient portal can lower the cost of health care. People consider information about lifestyles choices and data from devices to help monitor their health to be the least interesting information in a patient portal. The lack of interest to have data in a patient portal that is gathered from devices to monitor one’s health corresponds with findings from a previous study, which states that “tools alone are not enough” and engaging patients in the use of a patient portal requires a patient-centered approach [44].

Strengths and Limitations

This study is one of the first to investigate the health information needs of patients on conceptualizing a patient portal to access digital health data. In contrast to other studies in this field, which often focus on a functional patient portal, we had to start from the very beginning, as there was no functional patient portal available in Belgium at the time our survey was submitted [32,45-48]. In addition, the use of cultural probes and ideation techniques, frequently used in design and human computer interaction research, provides insights into the way people would like to interact with digital health data. It creates the opportunity to design a patient portal that considers the health information needs expressed by future users. Our survey was distributed across a large region in Belgium and included respondents of all age categories.

Despite our efforts to reduce selection bias to a minimum, we could not recruit participants from different ethnic backgrounds. Our respondents almost exclusively spoke Dutch. This is not in line with the demographics in Flanders, where a significant part of the population (8.1%) is of foreign origin. As almost 98% of respondents used the internet at least once a day, we could not find much evidence for the so-called “digital divide,” which is described in previous research as an important barrier in the implementation of a patient portal [49]. One potential explanation for this could be the Web-based nature of the survey and the fact that reminders to participate were mostly made with the help of social media and Web-based newsletters. This means our findings may not be representative of some groups in the Flemish population who rarely use the internet. Finally, our 2 statistical models had a low predictive power with an R^2 of 0.122 (predicted importance people attribute to obtaining health information when using a patient portal) and.106 (predicted expectations toward personal health care when using a patient portal). Although these R^2 values were low, this is not unusual in social research [50].

Conclusions

This study yields a range of relevant aspects to consider when designing a patient portal. First of all, people express the need for a patient portal and personal access to it. Second, people like to receive alerts or some form of communication to foster their health management. Finally, several patient characteristics influence people’s attitude toward a patient portal. As such, middle-aged people and those with a positive attitude toward shared decision making attach more importance to obtaining health information. People with a lower education level or with low health literacy expect an improvement in their health care by using a patient portal.

Acknowledgments

MyHealthData is a project cofunded by IMEC. Project partners are Mintlab, EBMPracticeNet, CEBAM, HealthConnect, Landsbond der Christelijke Mutualiteiten, and MindBytes, with project support from Agentschap Innoveren & Ondernemen. The authors gratefully acknowledge Martial Luyts and Anna Ivanova for their guidance with the statistical analyses.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Survey part 1 and part 2.

[PDF File (Adobe PDF File), 115KB - jmir_v20i10e269_app1.pdf]

Multimedia Appendix 2

Survey part 1 and part 2, Dutch translation.

[PDF File (Adobe PDF File), 760KB - jmir_v20i10e269_app2.pdf]

Multimedia Appendix 3

The Checklist for Reporting Results of Internet E-Surveys (CHERRIES).

[PDF File (Adobe PDF File), 48KB - jmir_v20i10e269_app3.pdf]

Multimedia Appendix 4

Qualitative study that forms the basis for the second part of the survey.

[PDF File (Adobe PDF File), 544KB - jmir_v20i10e269_app4.pdf]

Multimedia Appendix 5

Predictor variables for regression models.

[PDF File (Adobe PDF File), 48KB - jmir_v20i10e269_app5.pdf]

References

1. NHS. Equity and Excellence: Liberating the NHS. London: The Stationery Office/Tso; 2010.
2. 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 (DC): National Academies Press; 2001.
3. Barry MJ, Edgman-Levitan S. Shared decision making--pinnacle of patient-centered care. N Engl J Med 2012 Mar 01;366(9):780-781. [doi: 10.1056/NEJMp1109283] [Medline: 22375967]

4. 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;17(2):e40 [FREE Full text] [doi: 10.2196/jmir.3703] [Medline: 25707035]
5. 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]
6. Markle. NY, USA: Markle Connecting for Health; 2003 Jul 1. Personal Health Working Group Final Report URL: https://www.markle.org/publications/1429-personal-health-working-group-final-report [accessed 2018-06-24] [WebCite Cache ID 70Petps3V]
7. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc 2006 Mar;13(2):121-126 [FREE Full text] [doi: 10.1197/jamia.M2025] [Medline: 16357345]
8. Gheorghiu B, Hagens S. Use and Maturity of Electronic Patient Portals. Stud Health Technol Inform 2017;234:136-141. [Medline: 28186030]
9. Bates DW, Wells S. Personal health records and health care utilization. JAMA 2012 Nov 21;308(19):2034-2036. [doi: 10.1001/jama.2012.68169] [Medline: 23168828]
10. Otte-Trojel T, de Bont A, van de Klundert J, Rundall TG. Characteristics of patient portals developed in the context of health information exchanges: early policy effects of incentives in the meaningful use program in the United States. J Med Internet Res 2014;16(11):e258 [FREE Full text] [doi: 10.2196/jmir.3698] [Medline: 25447837]
11. Gerard M, Fossa A, Folcarelli PH, Walker J, Bell SK. What Patients Value About Reading Visit Notes: A Qualitative Inquiry of Patient Experiences With Their Health Information. J Med Internet Res 2017 Jul 14;19(7):e237 [FREE Full text] [doi: 10.2196/jmir.7212] [Medline: 28710055]
12. Black H, Gonzalez R, Priolo C, Schapira MM, Sonnad SS, Hanson CW, et al. True "meaningful use": technology meets both patient and provider needs. Am J Manag Care 2015 May 01;21(5):e329-e337 [FREE Full text] [Medline: 26167781]
13. 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;4(9):e006021 [FREE Full text] [doi: 10.1136/bmjopen-2014-006021] [Medline: 25200561]
14. 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]
15. Delbanco T, Walker J, Darer JD, Elmore JG, Feldman HJ, Leveille SG, et al. Open notes: doctors and patients signing on. Ann Intern Med 2010 Jul 20;153(2):121-125. [doi: 10.7326/0003-4819-153-2-201007200-00008] [Medline: 20643992]
16. Schnipper JL, Gandhi TK, Wald JS, Grant RW, Poon EG, Volk LA, et al. Effects of an online personal health record on medication accuracy and safety: a cluster-randomized trial. J Am Med Inform Assoc 2012;19(5):728-734 [FREE Full text] [doi: 10.1136/amiajnl-2011-000723] [Medline: 22556186]
17. Weingart SN, Hamrick HE, Tutkus S, Carbo A, Sands DZ, Tess A, et al. Medication safety messages for patients via the web portal: the MedCheck intervention. Int J Med Inform 2008 Mar;77(3):161-168. [doi: 10.1016/j.ijmedinf.2007.04.007] [Medline: 17581772]
18. Shimada SL, Allison JJ, Rosen AK, Feng H, Houston TK. Sustained Use of Patient Portal Features and Improvements in Diabetes Physiological Measures. J Med Internet Res 2016 Jul 01;18(7):e179 [FREE Full text] [doi: 10.2196/jmir.5663] [Medline: 27369696]
19. Irizarry T, DeVito DA, Curran CR. Patient Portals and Patient Engagement: A State of the Science Review. J Med Internet Res 2015;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
20. http://www.raosoft.com/samplesize.html. URL: http://www.raosoft.com/samplesize.html [accessed 2017-05-14] [WebCite Cache ID 6qS807qEy]
21. Patel VN, Abramson E, Edwards AM, Cheung MA, Dhopeshwarkar RV, Kaushal R. Consumer attitudes toward personal health records in a beacon community. Am J Manag Care 2011 Apr;17(4):e104-e120 [FREE Full text] [Medline: 21774099]
22. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
23. Hambleton R, Rodgers J. Item Bias Review. Practical Assessment, Research & Evaluation 1995:1-3.
24. Pouwer F, van der Ploeg H, Bramsen I. Dwalingen in de methodologie II. Bias door vragenlijsten. Nederlands tijdschrift voor geneeskunde 1998;142(27):1556-1558.
25. De Roeck D, Slegers K, Stappers P, Standaert A. Infusing digital creativity in design, the low fidelity way. In: Proceedings of the Chi Sparks. 2011 Jun 23 Presented at: Chi Sparks Nederland; 2011; Arnhem, The Netherlands.
26. Gaver B, Dunne T, Pacenti E. Design: Cultural probes. Interactions 1999 Jan;6(1):21-29. [doi: 10.1145/291224.291235]
27. Boehner K, Vertesi J, Sengers P, Dourish P. How HCI interprets the probes. 1240789: ACM; 2007 Presented at: The SIGCHI Conference on Human Factors in Computing Systems; 2007; San Jose, California, USA p. 1077-1086.
28. Vandenberghe B, Slegers K. Anthropomorphism as a Strategy to Engage End-Users in Health Data Ideation. 2971500: ACM; 2016 Presented at: The 9th Nordic Conference on Human-Computer Interaction; 2016; Gothenburg, Sweden p. 1-4.

29. Hype Stat. Knack.Be - Info URL: https://hypestat.com/info/knack.be [accessed 2018-05-06] [WebCite Cache ID 6zCszvzvE]
30. Norman G. Likert scales, levels of measurement and the "laws" of statistics. Adv in Health Sci Educ 2010 Feb 10;15(5):625-632. [doi: 10.1007/s10459-010-9222-y]
31. Wakefield BJ, Turvey CL, Nazi KM, Holman JE, Hogan TP, Shimada SL, et al. Psychometric Properties of Patient-Facing eHealth Evaluation Measures: Systematic Review and Analysis. J Med Internet Res 2017 Oct 11;19(10):e346 [FREE Full text] [doi: 10.2196/jmir.7638] [Medline: 29021128]
32. Alpert JM, Krist AH, Aycock RA, Kreps GL. Applying Multiple Methods to Comprehensively Evaluate a Patient Portal's Effectiveness to Convey Information to Patients. J Med Internet Res 2016 May;18(5):e112. [doi: 10.2196/jmir.5451] [Medline: 27188953]
33. Price-Haywood EG, Harden-Barrios J, Ulep R, Luo Q. eHealth Literacy: Patient Engagement in Identifying Strategies to Encourage Use of Patient Portals Among Older Adults. Popul Health Manag 2017 Dec;20(6):486-494. [doi: 10.1089/pop.2016.0164] [Medline: 28384076]
34. Coughlin SS, Stewart JL, Young L, Heboyan V, De Leo G. Health literacy and patient web portals. Int J Med Inform 2018 May;113:43-48. [doi: 10.1016/j.ijmedinf.2018.02.009] [Medline: 29602432]
35. Davis S, Roudsari A, Courtney KL. Designing Personal Health Record Technology for Shared Decision Making. Stud Health Technol Inform 2017;234:75-80. [Medline: 28186019]
36. Cutrona SL, Sreedhara M, Goff SL, Fisher LD, Preusse P, Jackson M, et al. Improving Rates of Influenza Vaccination Through Electronic Health Record Portal Messages, Interactive Voice Recognition Calls and Patient-Enabled Electronic Health Record Updates: Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2016 May 06;5(2):e56 [FREE Full text] [doi: 10.2196/resprot.5478] [Medline: 27153752]
37. Alpert JM, Krist AH, Aycock RA, Kreps GL. Designing User-Centric Patient Portals: Clinician and Patients' Uses and Gratifications. Telemed J E Health 2017 Dec;23(3):248-253 [FREE Full text] [doi: 10.1089/tmj.2016.0096] [Medline: 27333468]
38. Tambuyzer E, Van Audenhove C. Is perceived patient involvement in mental health care associated with satisfaction and empowerment? Health Expect 2015 Aug;18(4):516-526 [FREE Full text] [doi: 10.1111/hex.12052] [Medline: 23425015]
39. King G, Maxwell J, Karmali A, Hagens S, Pinto M, Williams L, et al. Connecting Families to Their Health Record and Care Team: The Use, Utility, and Impact of a Client/Family Health Portal at a Children's Rehabilitation Hospital. J Med Internet Res 2017 Apr 06;19(4):e97 [FREE Full text] [doi: 10.2196/jmir.6811] [Medline: 28385680]
40. 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]
41. de Lusignan S, Ross P, Shifrin M, Hercigonja-Szekeres M, Seroussi B. A comparison of approaches to providing patients access to summary care records across old and new europe: an exploration of facilitators and barriers to implementation. Stud Health Technol Inform 2013;192:397-401. [Medline: 23920584]
42. Vodicka E, Mejilla R, Leveille SG, Ralston JD, Darer JD, Delbanco T, et al. Online access to doctors' notes: patient concerns about privacy. J Med Internet Res 2013;15(9):e208 [FREE Full text] [doi: 10.2196/jmir.2670] [Medline: 24072335]
43. Clark SJ, Costello LE, Gebremariam A, Dombkowski KJ. A national survey of parent perspectives on use of patient portals for their children's health care. Appl Clin Inform 2015;6(1):110-119 [FREE Full text] [doi: 10.4338/ACI-2014-10-RA-0098] [Medline: 25848417]
44. Ryan BL, Brown JB, Terry A, Cejic S, Stewart M, Thind A. Implementing and Using a Patient Portal: A qualitative exploration of patient and provider perspectives on engaging patients. J Innov Health Inform 2016 Jul 04;23(2):848 [FREE Full text] [Medline: 27869582]
45. Otte-Trojel T, de Bont A, Rundall TG, van de Klundert J. What do we know about developing patient portals? a systematic literature review. J Am Med Inform Assoc 2016 Apr;23(e1):e162-e168. [doi: 10.1093/jamia/ocv114] [Medline: 26335985]
46. Ronda MCM, Dijkhorst-Oei L, Rutten GEHM. Patients' Experiences with and Attitudes towards a Diabetes Patient Web Portal. PLoS One 2015 Jun;10(6):e0129403 [FREE Full text] [doi: 10.1371/journal.pone.0129403] [Medline: 26086272]
47. Lazard AJ, Watkins I, Mackert MS, Xie B, Stephens KK, Shalev H. Design simplicity influences patient portal use: the role of aesthetic evaluations for technology acceptance. J Am Med Inform Assoc 2016 Apr;23(e1):e157-e161. [doi: 10.1093/jamia/ocv174] [Medline: 26635314]
48. Coughlin SS, Prochaska JJ, Williams LB, Besenyi GM, Heboyan V, Goggans DS, et al. Patient web portals, disease management, and primary prevention. Risk Manag Healthc Policy 2017;10:33-40 [FREE Full text] [doi: 10.2147/RMHP.S130431] [Medline: 28435342]
49. 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]
50. Moksony F. Small is beautiful. The use and interpretation of R 2 in social research. Szociologiai Szemle 1999:130-138.

Abbreviations

HITEC: Health Information Technology Evaluation Collaborative

---

Edited by G Eysenbach; submitted 26.11.17; peer-reviewed by T Irizarry, B Xie, C Kruse, R Pankomera; comments to author 15.03.18; revised version received 09.05.18; accepted 29.06.18; published 01.10.18.

Please cite as:
Van den Bulck SA, Hermens R, Slegers K, Vandenberghe B, Goderis G, Vankrunkelsven P
Designing a Patient Portal for Patient-Centered Care: Cross-Sectional Survey
J Med Internet Res 2018;20(10):e269
URL: http://www.jmir.org/2018/10/e269/
doi:10.2196/jmir.9497
PMID:

Original Paper

Responsiveness, Reliability, and Minimally Important and Minimal Detectable Changes of 3 Electronic Patient-Reported Outcome Measures for Low Back Pain: Validation Study

Robert Froud[1,2], PhD; Carol Fawkes[3], PhD; Jonathan Foss[1,4], PhD; Martin Underwood[1], MBChB, MD (UK); Dawn Carnes[3,5], PhD

[1]Clinical Trials Unit, Warwick Medical School, University of Warwick, Coventry, United Kingdom
[2]Institute of Health Sciences, Kristiania University College, Oslo, Norway
[3]Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom
[4]Department of Computer Science, University of Warwick, Coventry, United Kingdom
[5]Faculty of Health, University of Applied Sciences and the Arts, Western Switzerland, Switzerland

Corresponding Author:
Robert Froud, PhD
Clinical Trials Unit
Warwick Medical School
University of Warwick
Gibbet Hill Campus
Coventry, CV4 7AL
United Kingdom
Phone: 44 024 7657 4221
Email: r.froud@warwick.ac.uk

Abstract

Background: The Roland Morris Disability Questionnaire (RMDQ), visual analog scale (VAS) of pain intensity, and numerical rating scale (NRS) are among the most commonly used outcome measures in trials of interventions for low back pain. Their use in paper form is well established. Few data are available on the metric properties of electronic counterparts.

Objective: The goal of our research was to establish responsiveness, minimally important change (MIC) thresholds, reliability, and minimal detectable change at a 95% level (MDC_{95}) for electronic versions of the RMDQ, VAS, and NRS as delivered via iOS and Android apps and Web browser.

Methods: We recruited adults with low back pain who visited osteopaths. We invited participants to complete the eRMDQ, eVAS, and eNRS at baseline, 1 week, and 6 weeks along with a health transition question at 1 and 6 weeks. Data from participants reporting recovery were used in MIC and responsiveness analyses using receiver operator characteristic (ROC) curves and areas under the ROC curves (AUCs). Data from participants reporting stability were used for analyses of reliability (intraclass correlation coefficient [ICC] agreement) and MDC_{95}.

Results: We included 442 participants. At 1 and 6 weeks, ROC AUCs were 0.69 (95% CI 0.59 to 0.80) and 0.67 (95% CI 0.46 to 0.87) for the eRMDQ, 0.69 (95% CI 0.58 to 0.80) and 0.74 (95% CI 0.53 to 0.95) for the eVAS, and 0.73 (95% CI 0.66 to 0.80) and 0.81 (95% CI 0.69 to 0.92) for the eNRS, respectively. Associated MIC thresholds were estimated as 1 (0 to 2) and 2 (–1 to 5), 13 (9 to 17) and 7 (–12 to 26), and 2 (1 to 3) and 1 (0 to 2) points, respectively. Over a 1-week period in participants categorized as “stable” and “about the same” using the transition question, ICCs were 0.87 (95% CI 0.66 to 0.95) and 0.84 (95% CI 0.73 to 0.91) for the eRMDQ with MDC_{95} of 4 and 5, 0.31 (95% CI –0.25 to 0.71) and 0.61 (95% CI 0.36 to 0.77) for the eVAS with MDC_{95} of 39 and 34, and 0.52 (95% CI 0.14 to 0.77) to 0.67 (95% CI 0.51 to 0.78) with MDC_{95} of 4 and 3 for the eNRS.

Conclusions: The eRMDQ was reliable with borderline adequate responsiveness. The eNRS was responsive with borderline reliability. While the eVAS had adequate responsiveness, it did not have an attractive reliability profile. Thus, the eNRS might be preferred over the eVAS for measuring pain intensity. The observed electronic outcome measures’ metric properties are within the ranges of values reported in the literature for their paper counterparts and are adequate for measuring changes in a low back pain population.

(J Med Internet Res 2018;20(10):e272) doi:10.2196/jmir.9828

KEYWORDS

electronic patient-reported outcome measures; validation; responsiveness; reliability; minimally important change; minimal detectable change; Roland Morris Disability Questionnaire; visual analog scale; numerical rating scale

Introduction

Low back pain is a common and costly problem resulting in substantial personal, social, and economic burdens and is the number one cause of disability globally [1,2]. Low back pain is a symptom rather than a disease and most low back pain is nonspecific (ie, where no specific underlying cause has been identified, but where the term lacks formal definition and where definitions in trials have been diverse) [1,3]. The lifetime prevalence of low back pain is between 60% and 84% [4,5]. The global problem of low back pain is getting worse due to aging and increasing population size [6,7]. The number of clinical trials of interventions for low back pain has been increasing, with over 30 trials of interventions for low back pain now being published annually [8]. Patient-reported outcome measures (PROMs) in the form of paper questionnaires are typically used in these trials to judge the effectiveness of the health technology under investigation [8].

Disability and pain are by far the most commonly measured domains in trials of interventions for low back pain; each is measured at least twice as often as any other domain [8]. The visual analog scale (VAS) and numerical rating scale (NRS) are most commonly used for measuring pain intensity and the Roland Morris Disability Questionnaire (RMDQ) is most commonly used for measuring functional disability [8]. These are quasi-continuous measures where the relationship between the observed item responses and the unobserved latent variable is assumed to be consistent with a reflective conceptual framework [9]. There is evidence that paper forms of VAS and NRS have been in use since at least the early to mid-20th century, and the RMDQ has been used since 1983 [10-12].

The validity of a PROM is defined as “the degree to which an instrument truly measures the construct(s) it purports to measure” [13]. Several aspects that compose what we consider to constitute good development and validation of PROMs postdate the introduction of these particular instruments. Validation exercises have been performed retrospectively, results have accrued over time, and endorsement and use of the measures have survived the process [14-16]. Notwithstanding healthy academic debate, it is generally accepted that these outcome measures have reasonable face validity and content validity, and they have at times been considered the legacy gold standard for comparison for assessing the criterion/convergent validity of other instruments [17-19].

Measuring patient/participant change in health status using browser-based technology and mobile device technologies is a natural progression. Digital PROMs and ports of existing paper PROMs to digital media have become known as electronic patient-reported outcomes measures [20]. When migrating existing paper PROMs to electronic patient-reported outcome measures (ePROs), there are aspects relating to the metric validity of the instrument that may need to be reassessed. Some aspects of validity are clearly independent of whether the instrument is completed on paper or digitally—for example, the content wording (unless it is culturally or clinically out of date) and the extent to which this content is judged to appropriately span the domains of the health construct being measured (ie, content and face validity). However, other aspects of validity that relate directly to measurement performance should not be assumed to be unchanged.

For any instrument designed to measure change in a health construct, 2 properties are particularly relevant: reproducibility (ie, reliability) and responsiveness. Reliability is the extent to which the same results are obtained on repeated measures when no real change in health status has occurred [21,22]. An analogy using a bathroom scale is that it is desirable that the scale shows the same weight upon time-standardized daily measurement when there truly is no true change in a person’s weight; if this is the case, the scale may be said to be reliable. Conversely, responsiveness is analogous to the scale detecting an important change when one truly exists. As users’ physical interactions with ePRO versions of PROMs differs in fundamental respects from paper versions, we suggest that reassessing these 2 key change measurement properties is necessary before advocating their widespread use in health research.

In analyses of trials or evaluations of health interventions, using PROMs to decide when an individual participant has responded facilitates interpretation of intervention effect [23]. Responder analysis permits the number of improvements to simply be counted and compared by arm using several clear statistics. These are intuitive reporting methods, and there is consensus that back pain trials should incorporate these [23-25]. However, to be able to do this, it is necessary to know (1) the minimum thresholds considered important to an individual participant—the minimally important change (MIC)—and (2) what magnitudes of change can be detected beyond the inherent measurement error of the instrument—the minimal detectable change (MDC) [26,27]. These thresholds may be altered by the change in media from paper to digital and may also be population specific [28,29].

We aimed to determine reliability and responsiveness, MIC and MDC, for electronic versions of the VAS, RMDQ, and NRS as administered via Web browser and Android or iOS app to adults with low back pain who visit osteopaths.

Methods

Recruitment

We recruited adults with low back pain from osteopathic clinics in England and Wales. Participants were recruited by osteopaths on our behalf and provided with an enrollment code and instructions for installing the iOS or Android app (from the App

Store or Google Play) or completing the outcome measures using a Web browser.

We assumed an attrition rate of up to 70% and a recovery rate (ie, participants who indicate that they are much better or completely recovered using a health transition question) of over 90% in those with acute and subacute low back pain (ie, low back pain present for less than 3 months) [30]. Thus, for our responsiveness study, for which we required improved participants, we sought to recruit a minimum of 200 people with acute and subacute low back pain to ensure at least 50 eligible 6-week measurements. For people with chronic low back pain receiving manual therapy, we assumed up to the same rate of attrition but a lower rate of recovery (45%) [24]. For our test-retest study, we required stable participants who identified as remaining stable over a period of 1 week; thus, we sought to recruit 400 chronic patients to find 50 participants self-identifying as stable (ie, reporting no change on a health transition question). Participants were invited to complete the electronic versions of outcome measures at baseline, 1-week, and 6-week follow-up time points. Participants were offered a £5 (US $7) retail gift voucher for completing the outcome measures.

Software

We used Android and iOS apps and a Web app with an associated form builder that was developed by Clinvivo Ltd, a University of Warwick spin-out company [31]. The apps, which function identically across platforms, permitted PROMs to be typeset and then administered to patients securely on their own devices. Data in transit are encrypted using a Secure Sockets Layer, and data at rest are encrypted using a Rivest-Shamir-Aldeman and Advanced Encryption Standard encryption hybrid. At the end of the study period, data were encrypted using the open Pretty Good Privacy standard and transferred from Clinvivo to researchers. The iOS, Android, and Web apps sent data one way and did not receive or redisplay personal data. The platform presented an electronic version of the instrument and reminded participants to complete outstanding follow-up measurements, as appropriate. Off-line completion in apps was permitted in cases of interrupted connectivity, with submissions occurring upon restoration of connectivity. Reminders, which were received up to twice per follow-up measurement due, were sent directly to devices for app-enrolled participants and by email to Web-enrolled participants (up to 2 reminders).

Electronic Versions of Patient-Reported Outcome Measures

The VAS is a continuous scale running from 0 to 100 mm measuring current pain intensity [32]. It is the most commonly used outcome measure in trials of interventions for nonspecific low back pain overall [8]. Huskisson is commonly credited with its development in 1974; however, there is evidence that it was being used at least as far back as 1921 [11]. Intellectual property rights are in the public domain, and no permissions are required for use, reproductions, or modifications. Completion of the paper scale involves a person marking a line on the scale indicating their level of pain between 2 anchored scales that typically have wordings of “no pain” on the left (ie, 0 mm) and “worst possible pain” or “worst imaginable pain” on the right (ie, 100 mm) [33,34]. On paper, the distance of the marked line is then measured from the point of 0 pain and reported in mm. In migrating this to an electronic version (eVAS), we implemented a slider that could be dragged into position. We did not force the scale to render at 10 cm to allow for resizing to screens of different devices. Thus, we report scores in units rather than mm, where 1 unit is 1/100th of the scale (ie, where the pointer can be set at any one of 101 different positions) as rendered (Figure 1).

The RMDQ is a 24-item questionnaire measuring functional disability due to back pain that was developed in the early 1980s [10]. It is the most commonly used outcome measure in trials of interventions for low back pain overall [8]. The original paper version of the instrument is well established [35-38]. No permissions are required for its use, reproductions, or modifications [39]. Scores on the RMDQ range from 0 to 24, where higher scores indicate greater disability. Participants are given a statement with which they may indicate agreement by ticking a box. Participants are asked to tick statements that they feel describe them on that day and to leave blank boxes next to statements that they feel do not. The score is then the sum total of checked items. Our electronic (eRMDQ) migration is an exact copy using multiselect check boxes (Figure 2). One year into the research, we added a box stating “none of the above symptoms” for participants to confirm that none of the statements applied to them and to confirm 0 scores were genuine and not reflective of a skipped question.

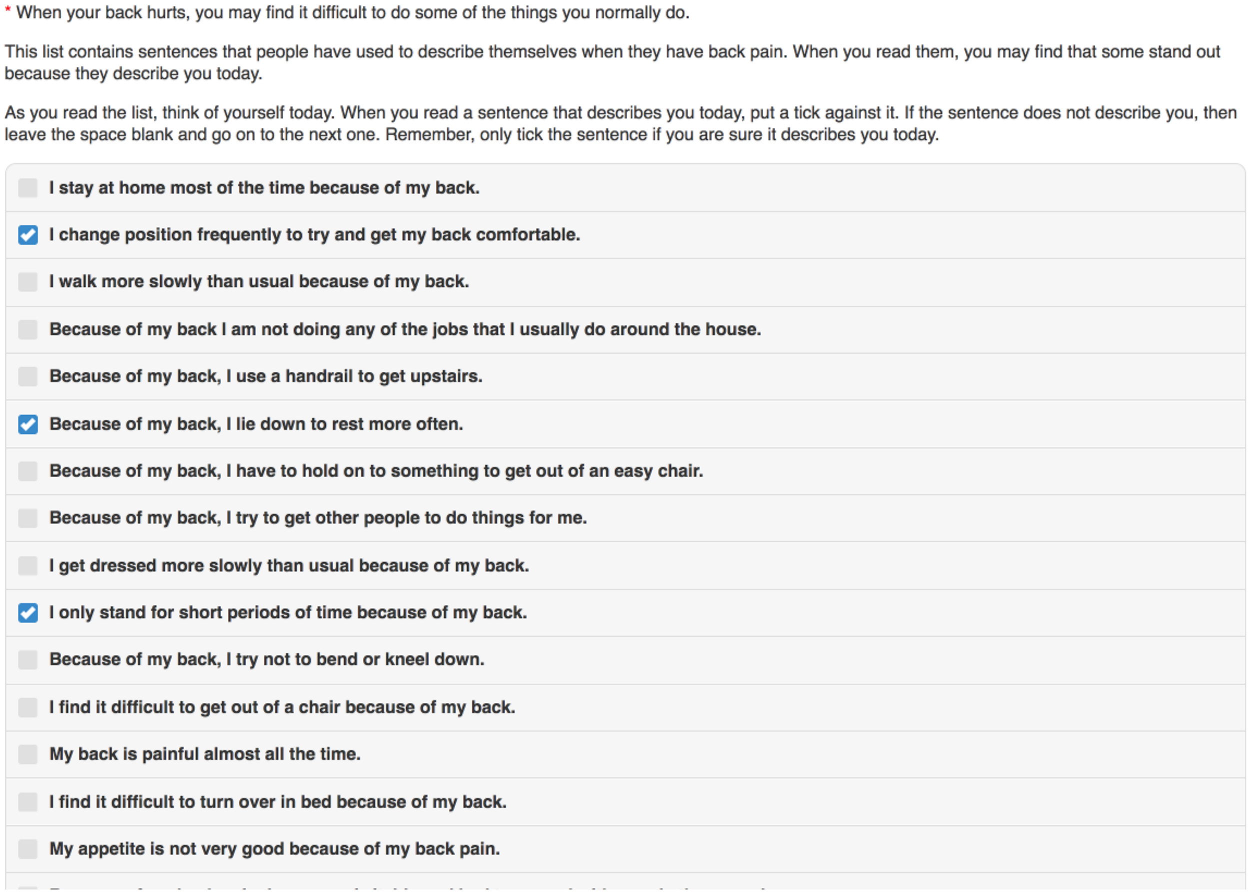
The NRS is an 11-point ordinal scale measuring current pain intensity [40,41]. Validation of the paper version is well established [41-43]. It is the fourth most commonly used outcome in trials of interventions for low back pain overall [8]. It is well established, with intellectual property rights in the public domain. Scores on the NRS range from 0, which typically is anchored “no pain” and 10, which typically is anchored “worst pain possible.” Our electronic (eNRS) migration is an exact copy with these anchor wordings (Figure 3). As the range of responses is exhaustive, completion of the scale was required for submission.

Figure 1. Electronic visual analog scale for pain intensity showing 63 units of pain intensity.

On average, how severe has your pain been over the past week?

(Please drag the blue slider)

Figure 2. Electronic Roland Morris Disability Scale showing a part score of 3 units.



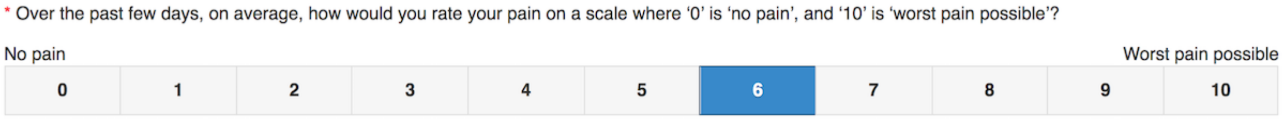
* When your back hurts, you may find it difficult to do some of the things you normally do.

This list contains sentences that people have used to describe themselves when they have back pain. When you read them, you may find that some stand out because they describe you today.

As you read the list, think of yourself today. When you read a sentence that describes you today, put a tick against it. If the sentence does not describe you, then leave the space blank and go on to the next one. Remember, only tick the sentence if you are sure it describes you today.

☐ I stay at home most of the time because of my back.
☑ I change position frequently to try and get my back comfortable.
☐ I walk more slowly than usual because of my back.
☐ Because of my back I am not doing any of the jobs that I usually do around the house.
☐ Because of my back, I use a handrail to get upstairs.
☑ Because of my back, I lie down to rest more often.
☐ Because of my back, I have to hold on to something to get out of an easy chair.
☐ Because of my back, I try to get other people to do things for me.
☐ I get dressed more slowly than usual because of my back.
☑ I only stand for short periods of time because of my back.
☐ Because of my back, I try not to bend or kneel down.
☐ I find it difficult to get out of a chair because of my back.
☐ My back is painful almost all the time.
☐ I find it difficult to turn over in bed because of my back.
☐ My appetite is not very good because of my back pain.

Figure 3. Electronic numerical rating scale for pain intensity showing a part score of 6 units.



* Over the past few days, on average, how would you rate your pain on a scale where ‘0’ is ‘no pain’, and ‘10’ is ‘worst pain possible’?

Participants were also asked to electronically complete a health transition question at 1- and 6-week follow-up time points. The transition question was a single question with the wording “Overall, how would you rate the change in your symptoms since beginning this study?” where the participant could respond on a 7-point scale [44]: 1–completely recovered, 2–much improved, 3–slightly improved, 4–no change, 5–slightly worsened, 6–much worsened, and 7–vastly worsened.

Assessment

We aimed to have 50 completed paired measurements in improving participants for responsiveness assessments and 50 completed test-retest measurements in stable participants. We defined improving participants a priori as participants who selected much improved or completely recovered using the transition question. Improving participant scores were used to assess responsiveness at 1 and 6 weeks. For our test-retest study, we defined stable participants a priori as those who select no change at 1 week, and in the case of having too few observations, a post hoc sensitivity analysis including those who selected either slightly worsened, no change, or slightly improved. This alternative “about the same” approach to marking stability has been used elsewhere [45]. Allowing 1 week is typical in low back pain test-retest studies; clinically, this is close enough for the people with chronic pain to remain stable but far enough apart that participants cannot easily recall their initial responses. It was anticipated that the chronic population would predominantly contribute participants to the test-retest study and improving participants would come from across all chronicity subpopulations.

Statistical Analyses

To measure responsiveness in a way that is consistent with the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) definition, we constructed receiver operator characteristic (ROC) curves for 1- and 6-week data using a dichotomized transition question as the external criterion [22]. The area under the ROC curve (AUC) is then a metric of responsiveness, accepting that the external criterion reasonably includes the construct of interest [46]. The approach has previously been used to quantify responsiveness across all 3 paper versions of instruments [47]. ROC AUCs of over 0.70 were considered to be adequate [9,48]. We dichotomized the transition question such that participants responding completely recovered and much improved were considered improved and all other responses were considered not improved.

We also used ROC curves and the transition question external criterion for 1- and 6-week data to quantify the MIC, which is defined as "the smallest [change] in score in the domain of interest which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient's management" (see note 1 in Multimedia Appendix 1) [43,49]. We used a MIC estimator based on the minimum sums of squares method, which consistently selects the cut-point closest to the top left corner of ROC space, as required when sensitivity and specificity are valued equally [50]. We calculated confidence intervals for MIC point estimates using bootstrapping [51].

To estimate reliability, we calculated intraclass correlation coefficients (ICCs) [52,53]. ICC values usually range from 0 to 1 [54]. ICC values above 0.75 may be interpreted as excellent agreement, values of 0.40 to 0.75 indicate poor to fair agreement, and values of below 0.40 indicate poor agreement [55]. We calculated the standard error of measurement [53]. We used this to estimate the minimal detectable change at the 95% level (MDC_{95}) (see notes 2 to 4 in Multimedia Appendix 1) [53,56,57].

Transition questions can be highly correlated with follow-up score rather than change [24,43,58]. Guyatt et al [58] assert that if a transition question is truly measuring change then a correlation between the baseline score and transition question and the follow-up score and transition question should ideally be present, equal, and opposite. In addition, they suggest that in a linear regression model with follow-up score entered as the initial explanatory variable, the baseline score should explain a significant proportion of the residual variance in the transition rating [58]. We performed Pearson correlations and fitted regression models to explore the degree to which the transition question measured change or simply reflected follow-up status. Log rank tests were used to assess significance of the addition of baseline score.

All analyses were performed using Stata version 14.2 (StataCorp LLC). The program rocmic was used to estimate MIC and the ROC AUC, which for ROC AUC uses the lroc program [51,59].

Power and Sample Size

With the notable exception of construct validity, sample sizes in validation studies generally are not calculated based on power to test hypotheses: the estimation of reliability and responsiveness parameters is focused on the extent to which the coefficients describing these parameters approach 1 (which would represent perfect reliability/responsiveness) rather than their difference from 0 or some other null value. Generally, a sample size of at least 50 participants is considered adequate for this purpose [9,60]. Assuming an ICC of 0.7, with 50 participants we would be able to estimate the ICC to within a 95% CI of +/–0.14. Alternatively, for an ICC of 0.8, we would be able to estimate to within a 95% CI of +/–0.10 [9]. For responsiveness, with 50 participants and assuming an AUC of 0.8 and equal numbers of cases and noncases, we would be able to estimate AUC to within a 95% CI of +/–0.12 [61].

As standard errors (SEs) for MIC estimates are not readily calculable, we used bootstrapping to generate SEs and 95% CIs [51,62]. Previous simulation work on the paper-based RMDQ in a similar population suggested that 2500 bootstrap samples was sufficient to ensure SE convergence [63]. To explore whether this is the case for the eRMDQ (and also whether it is an appropriate number of replications for the eNRS and eVAS), we simulated SEs by randomly sampling n observations (with replacement) from our dataset for an increasing number of n, where n is an integer, beginning at 20 and increasing by increments of 20, up to 6000 [62,64]. We then graphically assessed SE convergence and used the point of convergence to inform the number of bootstrap replications.

Data Exclusions, Assumptions, and Variations

Prior to the addition of the "none" box, we imputed 0 scores for all baseline submissions with no eRMDQ boxes ticked and assumed and imputed a 0 score for eRMDQ follow-up scores in the case that the baseline eRMDQ score was greater than 0 and a submission had been made for the follow-up period in question. When the eVAS rendered, it did so with the slider in the 0 position. In the case of a submission for an untouched eVAS, a score of 0 was assumed valid. The eNRS was a required response and necessitated a selection for submission.

As part of the basic demographic details collected, we included a list of presenting complaints, featuring low back pain among 15 other common musculoskeletal presentations and the opportunity to report a complaint not listed in a free-text box. The list of complaints was derived from earlier survey work developed as part of a national data collection initiative [65,66]. We excluded all cases where a participant had not checked the low back pain box (data from non-low back pain cases were used in unrelated research).

Ethics Approval

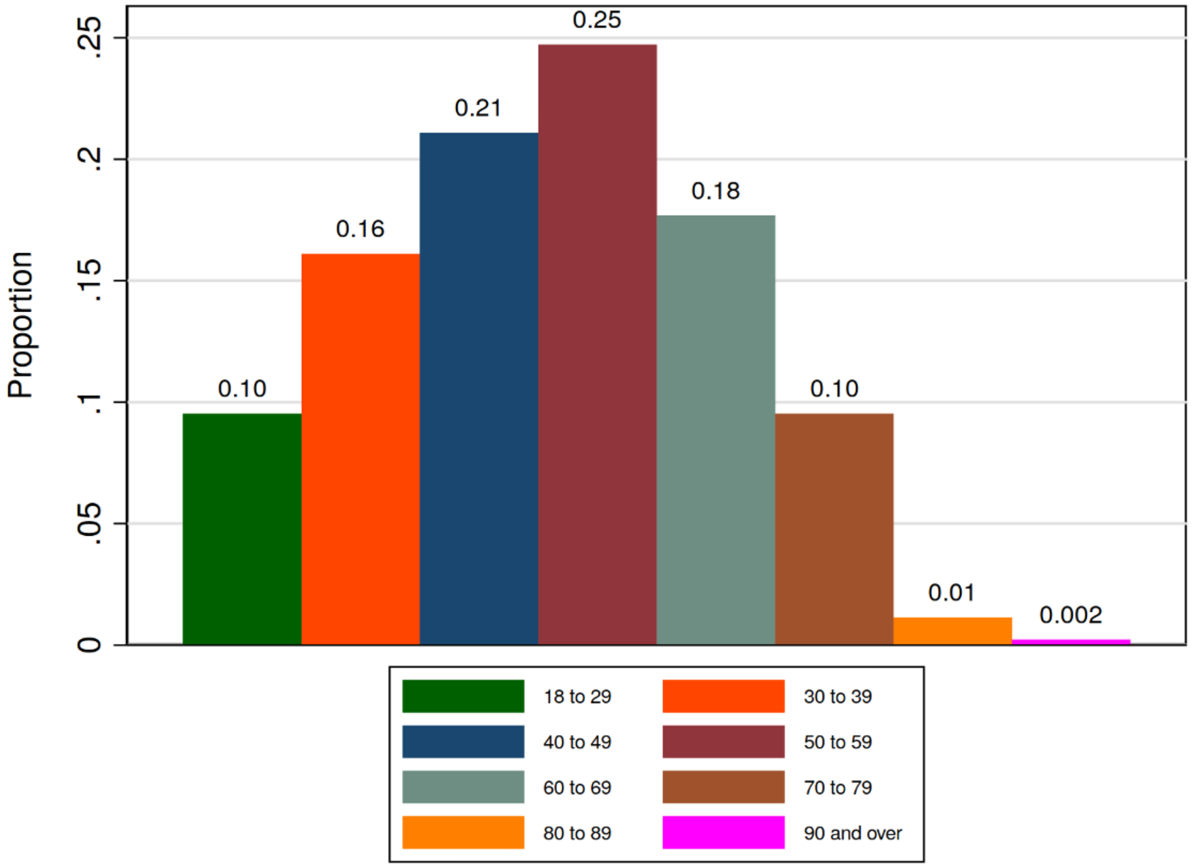
Ethics approval was obtained from the research ethics committee at Queen Mary University of London (QMERC2014/18).

Results

User Statistics and Demographics

We collected data from 575 people from 30 osteopathic clinics between July 15, 2014, and May 3, 2017. Of these, 442 (76.9%) reported low back pain as their main complaint. The average submission time for 1-week scores was 7.4 (SD 0.79) days after baseline. The average submission time for 6-week scores was 42.5 (SD 0.9) days after baseline. Of the participants, 60.4% (267/442) were female, 69.2% (306/442) identified as being in full or part-time employment, 1.1% (5/442) were long-term sick, 3.6% (16/442) identified as looking after home/family, 19.7% (87/442) were retired, 1.4% (6/442) were in full-time education, 2.9% (13/442) were unemployed, and 2.0% (9/442) selected other or preferred not to disclose. Figure 4 shows a histogram of patient-reported age at baseline.

Figure 4. Histogram of patient age at baseline.



We collected baseline eNRS data from 442 participants, and we collected baseline eVAS and eRMDQ data from 247 participants. One-week data were collected from 187 and 97 participants, respectively, and 6-week data were collected from 91 and 40 participants, respectively. Figure 5 shows the incidence of recovery in these groups. There was 1 missing data point for eNRS at baseline (0.2%) and 1 week (0.5%) for which we were unable to confirm cause. Table 1 summarizes ePRO submission scores using median and interquartile range and Table 2 summarizes recoveries and cumulative recoveries recorded using the transition question. Change scores (not shown) more closely followed normal distributions.

The addition of baseline score generally explained a significant proportion of the variance in the transition question over and above follow-up score. The transition question correlated with follow-up score but not with baseline score. Comprehensive results for the Guyatt analyses on the transition question's performance in measuring change are listed in note 2 in Multimedia Appendix 1.

Evaluation Outcomes

Graphically, SE convergence appeared to be asymptotically complete at around 5000 bootstrap replications (Figure 6); thus 5000 replications were used to generate confidence intervals for the MIC estimates in Table 3. Responsiveness point estimates (Table 3) were borderline adequate (AUC≈0.7) or above adequate for all instruments and time points. The AUC confidence interval for the RMDQ at 6 weeks spanned the null value (Table 3).

Using no change as a criterion for judging stability, we did not achieve our a priori threshold of 50 test-retest data points for comparison across any of the instruments. Of the people who said they had no change at 1 week, 65% (15/23) had chronic pain. Allowing slightly improved and slightly worsened to count as stable enabled us to achieve this threshold for the eNRS only. Of people who said they had no or slight change at 1 week, 63% (53/84) had chronic pain. Notwithstanding the lack of data, the eRMDQ reliability (agreement) was excellent using either analysis, with CIs spanning fair to excellent in both analyses (Table 4). For the eVAS per protocol analysis, the agreement was fair with CIs spanning poor to fair, and in the sensitivity analysis, the agreement was poor to fair with a CI range spanning poor to fair (Table 4). For the eNRS per protocol analysis, the agreement was poor to fair with a CI spanning poor to excellent, and for the sensitivity analysis, agreement was fair with a CI spanning poor to fair to excellent (Table 4).

Figure 5. Flowchart showing completion rates at 1 and 6 weeks, chronicity status, and the incidence of self-reported recovery using the health transition question for participants who also completed the electronic numerical rating scale, and electronic Roland Morris Disability Questionnaire, and electronic visual analog scale measurement.

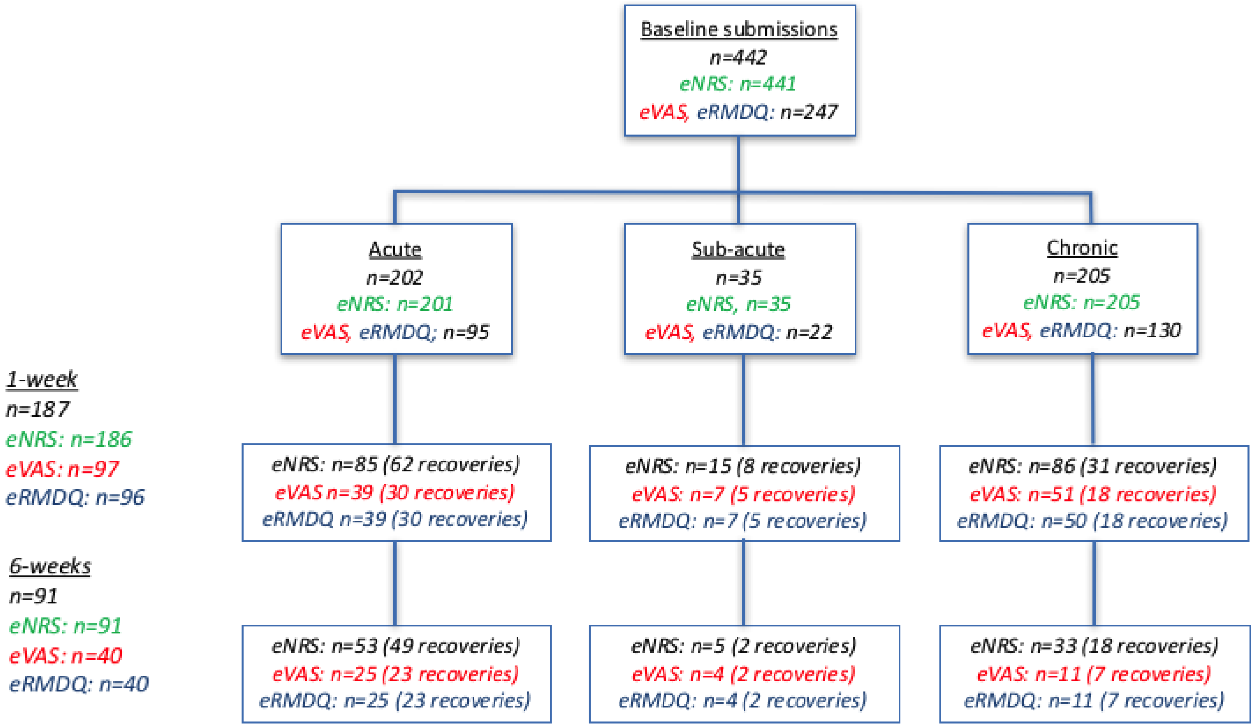


Table 1. Baseline, 1-week, and 6-week scores across the whole sample.

Score	Baseline		1 week		6 week	
	Median (IQR)	n[a]	Median (IQR)	n	Median (IQR)	n
eRMDQ[b]	4 (6)	247	2 (4)	96	2 (3.5)	40
eVAS[c]	41 (32)	247	24 (19)	97	19 (19)	40
eNRS[d]	5 (4)	441	3 (3)	186	2 (2)	91

[a] The number of received measurements at 1 week and at 6 weeks, respectively.

[b] eRMDQ: electronic Roland Morris Disability Questionnaire.

[c] eVAS: electronic visual analog scale.

[d] eNRS: electronic numerical rating scale.

Table 2. Recoveries and cumulative recoveries recorded using the transition question

Transition question	n	Recoveries, n (%)	Cumulative recoveries, n[a] (%)
1 week	187	101 (54)	101 (23)
6 weeks	91	69 (76)	170 (38)

[a] Where the frequency of cumulative recoveries are shown as a proportion of all 442 baseline participants.

Figure 6. Graphs showing minimally important change bootstrap standard error convergence from simulations with increasing replication numbers. MIC: minimally important change, NRS: numerical rating scale, RMDQ: Roland Morris Disability Questionnaire, VAS: visual analog scale.

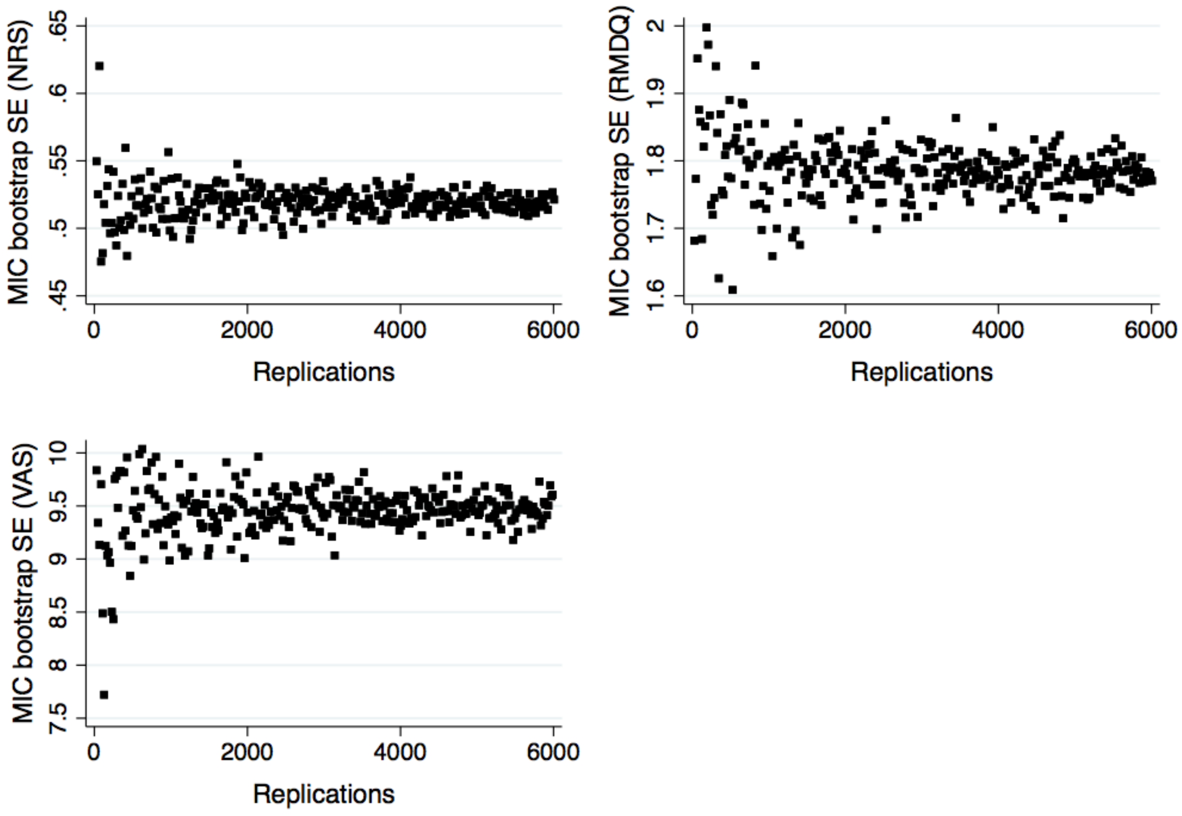


Table 3. Responsiveness and minimally important change by instrument and 1-week and 6-week follow-up time periods.

Instrument and time period	Receiver operator characteristic AUC[a]	95% CI	n[b]	Minimally important change points/ eVAS[c] units (% of baseline score)	95% CI
eRMDQ[d]					
1 week	0.69	0.59 to 0.80	96	1 (19)	0 to 2
6 weeks	0.67	0.46 to 0.87	40	2 (38)	–1 to 5
eVAS					
1 week	0.69	0.58 to 0.80	93	13 (32)	9 to 17
6 weeks	0.74	0.53 to 0.95	40	7 (17)	–12 to 26
eNRS[e]					
1 week	0.73	0.66 to 0.80	185	2 (43)	1 to 3
6 weeks	0.81	0.69 to 0.92	91	1 (21)	0 to 2

[a] AUC: area under the curve.

[b] The number of change scores available (ie, from available pairs of measurements at baseline and follow-up time point) at 1 week and 6 weeks, respectively.

[c] eVAS: electronic visual analog scale.

[d] eRMDQ: electronic Roland Morris Disability Questionnaire.

[e] eNRS: electronic numerical rating scale.

Table 4. Intraclass correlation coefficients from test-retest study in a per protocol stable sample and a pseudo-stable sample with associated minimal detectable change thresholds.

Instrument and condition	n[a]	Intraclass correlation coefficient$_{agreement}$	95% CI	MDC$_{95}$[b] points/eVAS[c] units
eRMDQ[d]				
Per protocol	15	0.87	0.66 to 0.95	4
Allowing slight change	43	0.84	0.73 to 0.91	5
eVAS				
Per protocol	15	0.31	–0.25 to 0.71	39
Allowing slight change	43	0.61	0.36 to 0.77	34
eNRS[d]				
Per protocol	22	0.52	0.14 to 0.77	4
Allowing slight change	83	0.67	0.51 to 0.78	3

[a] The number of cases satisfying the condition for analysis as a stable case.

[b] MDC$_{95}$: minimal detectable change at the 95% level.

[c] eVAS: electronic visual analog scale.

[d] eRMDQ: electronic Roland Morris Disability Questionnaire.

[e] eNRS: electronic numerical rating scale.

Discussion

Principal Findings

The results suggest that the eRMDQ had borderline adequate responsiveness and excellent reliability. Conversely, the eNRS had relatively good responsiveness at 6 weeks but borderline adequate reliability. The eNRS outperformed the eVAS, which had adequate responsiveness but relatively poor reliability. As test-retest numbers were few, eVAS CIs spanned poor to excellent, and thus further investigation is warranted. While exploring use by age was not a specific study objective, we note the results indicate encouraging use by older people from this population.

Comparison With Prior Work

Across acute and chronic back pain populations there has been like-for-like evaluation (ie, using similar and directly comparable methods) of the properties of paper versions of the outcome measures explored. ROC AUC for the RMDQ ranges from 0.64 to 0.93 [45,47,67-75]. ROC AUC for the NRS ranges from 0.67 to 0.93 [41,42,47,67,75,76]. ROC AUC for the VAS ranges from 0.71 to 0.93 [47,72,77-79]. Our results are within these ranges at 6 weeks for all instruments and for all but our eVAS instrument at 1 week, where our point estimate approaches the lower border of the range. Our eVAS data are nevertheless consistent with the range (ie, insofar as the upper CI overlaps). Estimates of ROC AUC for the VAS are fewer in the literature, which might explain why the range of reported results is narrower than it is for the RMDQ and NRS.

MIC thresholds for RMDQ ranged between 1.5 and 5.0 [21,24,35,67,68,72,75,80-83], for the NRS between 1.5 and 4.0 [41-43,67,75,81,84], and for the VAS between 15 and 28 mm [72]. Our absolute MIC thresholds are comparable but are toward the lower side of this range. MIC estimates are known to increase with baseline severity, and relatively low baseline scores likely explain our relatively low thresholds [68,75,81,84]. However, MIC thresholds in our results, expressed as percentage change from baseline, average 28% across all 3 instruments and all time points. This is consistent with the suggestion of Ostelo et al [29] (following their review of MIC and MDC literature) for using an improvement of between 20% and 30% of baseline score for the RMDQ, NRS, and VAS as a MIC threshold. We emphasize that the MIC thresholds relate to the degree of change that may be considered important for an individual and not what degree of difference may be considered important at a population level [27,85,86]. We note that the 2 negative CIs imply consistency of the data, with the true MIC thresholds being in the opposite direction of improvement (ie, a slight deterioration). This is likely an artifact of low power, and we suggest using inflated sample sizes for future studies based on the bootstrapped standard error observations.

Reported ICC estimates for the RMDQ have ranged from 0.42 to 0.95 [45,67,81,87] and for the NRS from 0.92 to 0.98 [67,81], and an estimate for the VAS of 0.71 has been reported [88]. Our results are within the ranges reported, but our ICC point estimate for the eVAS is lower than the reported paper VAS estimate. It is conceivable that rendering the eVAS slider in a 0 position might lead to additional variance in the case that the outcome is overlooked (ie, leading to a comparatively lower ICC), and future research might explore whether a touch to confirm 0 design is acceptable to users. We also note that some of the ICC values in the literature ranges may have been derived from ICCs for consistency rather than agreement; this is a practice known to exist (although it is not always clear which approach has been used) and known to overestimate reliability [53].

MDC$_{95}$ estimates reported (or in the case of the NRS only, either reported or calculated from reported standard error of measurements) have ranged from 5.0 to 12.1 for the RMDQ [21,24,35,45,56,67,81,83], from 2.4 to 11 (ie, almost the full width of the scale) for the NRS [41,45,67,81,84], and from 21.0 to 33.5 for the VAS [79,88,89]. Our estimates are slightly better than average for the RMDQ, toward the lower end of the range for the NRS, and comparable to the available estimates for the VAS.

In terms of comparison to studies assessing these instruments as ePROs, Bird et al [90] conducted a test-retest study among 22 healthy adults of the VAS administered on a tablet and found ICCs of 0.90 (0.82 to 0.95) as compared to 0.96 (0.92 to 0.98) in a paper version that participants completed simultaneously. It is difficult to compare the results with this study, as the time between test and retest was less than 30 minutes. A much shorter period between test and retest might be appropriate in some populations (eg, where change in acute pain must be measured over short spaces of time). In these cases, participants may be more prone to panel conditioning, where the second response is affected by recall of the first response [91]. For back pain, most interventions focus on chronic pain and longer time periods. When exploring reliability of low back pain outcome measures, a 1-week gap between test and retest is typical. Bijur et al [92] and Gallagher et al [93] have used small time frames between tests on a paper-based VAS in acute pain populations and demonstrate similarly high ICCs of 0.97 (0.96 to 0.98) and 0.99 (0.989 to 0.992), respectively. Also of relevance but not directly comparable is work by Bishop et al [94], who administered the RMDQ on paper and online and constructed limits of agreement, demonstrating equivalence with a score difference of only 0.03 points and a Bland-Altman range of –2.77 to 2.83.

Finally, we note that the distribution of the user age of the health outcomes app in this population appears to be higher than the age of health app users [95].

Implications

None of our results differs materially from ranges observed in population-similar and methodologically alike studies of paper counterparts. There is thus some suggestion that the ePROs under evaluation are suitable substitutes for PROMs for measuring change in low back pain. The eNRS outperformed the eVAS in terms of responsiveness and reliability. As such, we suggest the eNRS might be preferred over the eVAS for the measurement of low back pain intensity, but we caution that subsequent confirmatory research is warranted.

Limitations

The principal limitation is that in several cases we had small sample sizes. We had intended to recruit sufficient numbers to have at least 50 people for each assessment, in line with recommendations, but we failed to meet these targets, mainly as we underestimated the incidence of stability, although we also underestimated attrition [9]. There were high rates of improvement in people receiving treatment, and this is a hazard of nesting a test-retest design within a protocol where participants are receiving routine clinical treatment. This was of consequence in the eRMDQ responsiveness analysis, where the data are consistent with a null population parameter and thus 6-week responsiveness of the eRMDQ requires confirmation in a larger sample. Having too few data has greater implications for the test-retest assessment of the VAS where the CIs span coefficient values that can be interpreted at their extremes as either poor or excellent. It is less of an issue for the eRMDQ because while the numbers are low and lower at 1 and 6 weeks, respectively, the stronger signal combined with boundary proximity leads to narrower and more useful CIs.

It is not ideal that we permitted slightly worse and slightly improved categories to indicate stability in our test-retest, although we note a similar approach has been observed previously [45]. Further, this was a post hoc decision taken in light of having too few observations to use our more stringent a priori criterion of including only those reporting no change. The results using our a priori approach but with few observations are offered as sensitivity analyses that may provide useful comparison.

Having relatively few observations also meant that we were unable to explore differences by platform (ie, iOS, Android, and Web browser) or explore MIC as a function of baseline score (eg, stratifying by number in category of severity) or separately by chronicity, which may have been useful and allowed us to explore any differences in these metrics by chronicity. Thus, our focus here is pragmatic and results are generalizable to the population of adults with low back pain who consult osteopaths, notwithstanding chronicity.

We recorded in our database only the summed eRMDQ score rather than individual responses. Had we retained detail of individual response profiles of the eRMDQ, we could have also calculated internal consistency (as well as aspects of modern test theory: Rasch analysis to examine item performance or factor analysis to explore data dimensionality). Whereas COSMIN conflate internal consistency with reliability in their taxonomy [22,96], we consider internal consistency to be an indication of the unidimensionality of a scale and of item redundancy rather than the degree to which a scale is free from measurement error. As such, and with respect to the reliability definition, we preferred to consider it separately. We had not immediately considered that the media used for completion might affect internal consistency or item functioning of a scale. On reflection, however, we think that it is conceivable that presenting the scale digitally may alter the way patients respond in such a way that these could be affected. Additionally, there may be self-selection effects of those more familiar with digital media joining the study, and this may be a factor that could be confounded with how a person responds.

It is not ideal that our transition question correlates with follow-up score but not with baseline score. This is emerging to be the case generally and is not something particular to evaluating electronic outcome measures [24,43,58]. This emergence in our view raises the more general question of whether it is appropriate to use transition questions at all to evaluate change in outcome measures. Apart from being overly driven by follow-up score, the assumption that the transition question is sufficiently driven by the same latent construct as the PROM, to the extent that it may be considered a gold standard, may be unrealistic. We have previously explored what people think about when they complete the transition question and what they think about when they complete the paper RMDQ version, and we found discordance [97]. Pain appears to be a greater driver of the transition question, and the wording of the transition question (ie, attempting to place focus specifically on function or an explicit domain) does not appear to matter. In our study, we used the term symptoms. However, in the case that the suggestion arising from our previous research is incorrect, using a generic wording in the transition question might have the advantage of not favoring any one ePRO over another but the disadvantage of disassociating the transition question from any specific latent health construct. Use of a generically worded transition question would then introduce some information bias—for example, if people systematically attend more to a particular domain upon reading the word symptoms. We caution that the logic of the typically taken approach of using one outcome measure as a proxy gold standard of recovery and then using this proxy to judge domain-specific responsiveness and MIC thresholds in another may be questionable where there is domain mismatch.

There was a small amount of missing data at baseline and 1 week (a person in each case), which should have been impossible because a selection on the eNRS was a required response. We are uncertain of the cause but we suspect this might have been due to use of an obscure and/or obsolete browser.

This research was conducted solely in private care and people who pay to see osteopaths may differ from those attending publicly funded health care, as is more routinely the case in health services research. We note a lower than typical baseline severity (as compared to clinical trials) and thus some caution is indicated before generalizing to typical trial populations. Finally, our focus here was on the most commonly used domains and outcome measures in trials. The VAS is most commonly used overall (pain), RMDQ second most common (disability), and the NRS fourth most commonly used (pain). We did not include the third most commonly used outcome, the Oswestry Disability Questionnaire, which also measures disability [8]. Unlike the VAS and NRS, which are both single-item instruments, including two full disability questionnaires risked being unduly burdensome for participants. Qualitative work suggests that participants would prefer to spend only 5 to 10 minutes completing ePROs [98,99]. Including a direct comparison with paper versions would have permitted direct exploration of criterion validity; however, this approach would likely have been affected by panel condition and further added to participant burden.

Recommendations for Future Research

Sampling stable participants from people receiving routine clinical treatment allows the nesting of a test-retest design and makes for an efficient design. However, it produces some challenges for achieving sufficient recruitment over a realistic time period. It assumes that the transition question classification of unchanged is valid. As data suggest that transition question is driven more by follow-up state than change, the approach has some limitations. It would be scientifically preferable that test-retest studies are conducted within untreated populations. However, this has ethical and practical implications. When planning to nest a test-retest design within any treatment-containing protocol, based on rates observed in this study (using the lower eNRS no chance incidence), we recommend planning a study that is around 3 times larger (ie, seeking approximately 1200 people to obtain 50 stable participants). For study of responsiveness alone, about 250 participants should be sufficient to achieve 50 improvements at 6 weeks. The most extreme MIC threshold we estimated was 7 units (–12 to 26) for the eVAS at 6 weeks. This is lower than has been noted in studies of paper counterparts. Assuming the point estimate is representative of the population parameter, approximately 300 participants would be required to power a study to confirm the finding.

Retaining data at item level in future studies will permit more sophisticated analytics. There may need to be a cultural change as we transition from paper to digital measurement. The ability to more easily retain greater data resolution is a clear advantage of digital measurement and one that would be sensible to exploit. Further advantages in terms of cost, logistics, form validation, reminders, time logging, environmental factors, and reach are undeniable and, in our view, make electronic health measurement very attractive. More generally, routine outcome measurement in clinical practice may facilitate so-called learning health care systems and should be a shared goal of stakeholders across health care [100,101]. To achieve this, greater collaboration may be needed between clinicians, informatics specialists, and policy makers. We also encourage further metric testing of electronic versions of these and other legacy PROMs so that results may inform health services researchers and clinicians' choices of measure.

Conclusion

Each of the electronic outcome measures has metric properties that do not materially differ from values reported in the literature for their paper counterparts. A possible exception may be the reliability of the eVAS, for which there is insufficient existing research to make useful comparisons between paper and digital versions. The eRMDQ is adequate for measuring back-related disability, and the eNRS is adequate for measuring pain intensity. The eNRS should be preferred over the eVAS for the measurement of pain intensity.

Acknowledgments

Part of RF and JF's time on the study was funded by the Warwick Impact Fund, which administers the Higher Education Innovation Funding grant. The study was sponsored by University of Warwick but conducted from Queen Mary University of London, which sponsored the remainder of CF's doctoral research. Neither the sponsor nor funder had any involvement in study design, analysis, or reporting of results.

Authors' Contributions

RF and MU conceived of the study and applied for and were awarded the funding to do the study. CF undertook the day-to-day management and submitted the documents for consideration by the Queen Mary University of London ethics committee. JF and RF were responsible for data management. RF performed all analyses. RF wrote the first draft of the paper. All authors commented on and approved the manuscript.

Conflicts of Interest

RF, MU, and JF are directors and shareholders of Clinvivo Ltd, the University of Warwick spin-out company that provided the software for data collection in this study. The Higher Education Innovation Funding grant paid for the development of intellectual property licensed to Clinvivo and used in this study and also paid for UK retail vouchers used as incentives to recruit participants into the study. RF and DC are nonpracticing osteopaths; CF is a practicing osteopath. MU was chair of the National Institute for Health and Care Excellence accreditation advisory committee, for which he received a fee, until March 2017. MU is chief investigator or co-investigator on multiple previous and current research grants from the UK National Institute for Health Research (NIHR), Arthritis Research UK, Arthritis Australia, and the Australian National Health and Medical Research Council. He has received travel expenses for speaking at conferences from professional organizations hosting the conferences. He is an editor of the NIHR journal series for which he receives a fee. RF and MU have published multiple papers on chronic pain, some of which are referenced in this paper. RF, MU, and JF are part of an academic partnership with Serco Ltd related to return-to-work initiatives.

Multimedia Appendix 1

Technical notes and extended technical results for Guyatt analyses.

[PDF File (Adobe PDF File), 18KB - jmir_v20i10e272_app1.pdf]

References

1. Hartvigsen J, Hancock MJ, Kongsted A, Louw Q, Ferreira ML, Genevay S, Lancet Low Back Pain Series Working Group. What low back pain is and why we need to pay attention. Lancet 2018 Mar 20:1. [doi: 10.1016/S0140-6736(18)30480-X] [Medline: 29573870]
2. Froud R, Patterson S, Eldridge S, Seale C, Pincus T, Rajendran D, et al. A systematic review and meta-synthesis of the impact of low back pain on people's lives. BMC Musculoskelet Disord 2014 Feb 21;15:50 [FREE Full text] [doi: 10.1186/1471-2474-15-50] [Medline: 24559519]
3. Amundsen PA, Evans DW, Rajendran D, Bright P, Bjørkli T, Eldridge S, et al. Inclusion and exclusion criteria used in non-specific low back pain trials: a review of randomised controlled trials published between 2006 and 2012. BMC Musculoskelet Disord 2018 Apr 12;19(1):113 [FREE Full text] [doi: 10.1186/s12891-018-2034-6] [Medline: 29650015]
4. Airaksinen O, Brox JI, Cedraschi C, Hildebrandt J, Klaber-Moffett J, Kovacs F, COST B13 Working Group on Guidelines for Chronic Low Back Pain. European guidelines for the management of chronic nonspecific low back pain. Eur Spine J 2006 Mar;15 Suppl 2:S192-S300 [FREE Full text] [doi: 10.1007/s00586-006-1072-1] [Medline: 16550448]
5. Maher C, Underwood M, Buchbinder R. Non-specific low back pain. Lancet 2017 Feb 18;389(10070):736-747. [doi: 10.1016/S0140-6736(16)30970-9] [Medline: 27745712]
6. GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2017 Sep 16;390(10100):1211-1259 [FREE Full text] [doi: 10.1016/S0140-6736(17)32154-2] [Medline: 28919117]
7. Clark S, Horton R. Low back pain: a major global challenge. Lancet 2018 Mar 20:1. [doi: 10.1016/S0140-6736(18)30725-6] [Medline: 29573869]
8. Froud R, Patel S, Rajendran D, Bright P, Bjørkli T, Buchbinder R, et al. A systematic review of outcome measures use, analytical approaches, reporting methods, and publication volume by year in low back pain trials published between 1980 and 2012: respice, adspice, et prospice. PLoS One 2016;11(10):e0164573 [FREE Full text] [doi: 10.1371/journal.pone.0164573] [Medline: 27776141]
9. de Vet H, Terwee C, Mokkink L, Knol D. Measurement in Medicine. Cambridge: Cambridge University Press; 2011.
10. Roland M, Morris R. A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. Spine (Phila Pa 1976) 1983 Mar;8(2):141-144. [Medline: 6222486]
11. Hayes M, Patterson D. Experimental development of the graphic rating method. Psychol Bull 1921;18:98-99.
12. Leavell H. Contributions of the social sciences to the solution of health problems. N Engl J Med 1952 Dec 04;247(23):885-897. [doi: 10.1056/NEJM195212042472305] [Medline: 13002642]
13. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010 Jul;63(7):737-745. [doi: 10.1016/j.jclinepi.2010.02.006] [Medline: 20494804]

14. Deyo RA, Battie M, Beurskens AJ, Bombardier C, Croft P, Koes B, et al. Outcome measures for low back pain research: a proposal for standardized use. Spine (Phila Pa 1976) 1998 Sep 15;23(18):2003-2013. [Medline: 9779535]
15. Dworkin R, Turk D, Farrar J, Haythornthwaite J, Jensen M, Katz N, IMMPACT. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005 Jan;113(1-2):9-19. [doi: 10.1016/j.pain.2004.09.012] [Medline: 15621359]
16. Deyo RA, Dworkin SF, Amtmann D, Andersson G, Borenstein D, Carragee E, et al. Report of the NIH task force on research standards for chronic low back pain. Spine J 2014 Aug 01;14(8):1375-1391. [doi: 10.1016/j.spinee.2014.05.002] [Medline: 24950669]
17. Kopec JA, Esdaile JM, Abrahamowicz M, Abenhaim L, Wood-Dauphinee S, Lamping DL, et al. The Quebec Back Pain Disability Scale. Measurement properties. Spine (Phila Pa 1976) 1995 Feb 01;20(3):341-352. [Medline: 7732471]
18. Tan G, Jensen MP, Thornby JI, Shanti BF. Validation of the Brief Pain Inventory for chronic nonmalignant pain. J Pain 2004 Mar;5(2):133-137. [doi: 10.1016/j.jpain.2003.12.005] [Medline: 15042521]
19. Nishiwaki M, Takayama M, Yajima H, Nasu M, Kong J, Takakura N. The Japanese version of the Massachusetts General Hospital acupuncture sensation scale: a validation study. Evid Based Complement Alternat Med 2017;2017:7093967 [FREE Full text] [doi: 10.1155/2017/7093967] [Medline: 28676831]
20. Wallwiener M, Matthies L, Simoes E, Keilmann L, Hartkopf AD, Sokolov AN, et al. Reliability of an e-PRO tool of EORTC QLQ-C30 for measurement of health-related quality of life in patients with breast cancer: prospective randomized trial. J Med Internet Res 2017 Sep 14;19(9):e322 [FREE Full text] [doi: 10.2196/jmir.8210] [Medline: 28912116]
21. de Vet HC. Reproducibility and responsiveness of evaluative outcome measures. Int J Technol Assess Healthc 2001;17(4):479-487. [Medline: 11758292]
22. Mokkink L, Terwee C, Patrick D, Alonso J, Stratford P, Knol D, et al. COSMIN Checklist Manual version 9. 2012. URL: http://www.cosmin.nl/images/upload/files/COSMIN%20checklist%20manual%20v9.pdf [accessed 2018-08-29] [WebCite Cache ID 722fv3jFU]
23. Froud R, Underwood M, Carnes D, Eldridge S. Clinicians' perceptions of reporting methods for back pain trials: a qualitative study. Br J Gen Pract 2012 Mar;62(596):e151-e159 [FREE Full text] [doi: 10.3399/bjgp12X630034] [Medline: 22429424]
24. Froud R, Eldridge S, Lall R, Underwood M. Estimating the number needed to treat from continuous outcomes in randomised controlled trials: methodological challenges and worked example using data from the UK Back Pain Exercise and Manipulation (BEAM) trial. BMC Med Res Methodol 2009 Jun 11;9:35 [FREE Full text] [doi: 10.1186/1471-2288-9-35] [Medline: 19519911]
25. Froud R, Eldridge S, Kovacs F, Breen A, Bolton J, Dunn K, et al. Reporting outcomes of back pain trials: a modified Delphi study. Eur J Pain 2011 Nov;15(10):1068-1074. [doi: 10.1016/j.ejpain.2011.04.015] [Medline: 21596600]
26. de Vet HC, Terwee C, Ostelo R, Beckerman H, Knol D, Bouter L. Minimal changes in health status questionnaires: distinction between minimally detectable change and minimally important change. Health Qual Life Outcomes 2006;4(54):1. [Medline: 16925807]
27. Froud R, Underwood M, Eldridge S. Improving the reporting and interpretation of clinical trial outcomes. Br J Gen Pract 2012 Oct;62(603):e729-e731 [FREE Full text] [doi: 10.3399/bjgp12X657008] [Medline: 23265234]
28. Henschke N, van Enst A, Froud R, Ostelo RWG. Responder analyses in randomised controlled trials for chronic low back pain: an overview of currently used methods. Eur Spine J 2014 Apr;23(4):772-778 [FREE Full text] [doi: 10.1007/s00586-013-3155-0] [Medline: 24419902]
29. Ostelo R, Deyo R, Stratford P, Waddell G, Croft P, Von Korff M, et al. Interpreting change scores for pain and functional status in low back pain: towards international consensus regarding minimal important change. Spine 2008;33(1):90-94. [Medline: 18165753]
30. van Tulder M, Becker A, Bekkering T, Breen A, del Real MTG, Hutchinson A, COST B13 Working Group on Guidelines for the Management of Acute Low Back Pain in Primary Care. Chapter 3. European guidelines for the management of acute nonspecific low back pain in primary care. Eur Spine J 2006 Mar;15 Suppl 2:S169-S191 [FREE Full text] [doi: 10.1007/s00586-006-1071-2] [Medline: 16550447]
31. Clinvivo. 2018. URL: http://www.clinvivo.com/ [accessed 2018-01-12] [WebCite Cache ID 6wPydLMrk]
32. Huskisson E. Measurement of pain. Lancet 1974 Nov 09;2(7889):1127-1131. [Medline: 4139420]
33. Hägg O, Fritzell P, Nordwall A, Swedish Lumbar Spine Study Group. The clinical importance of changes in outcome scores after treatment for chronic low back pain. Eur Spine J 2003 Feb;12(1):12-20. [doi: 10.1007/s00586-002-0464-0] [Medline: 12592542]
34. Scott J. Vertical or horizontal visual analogue scales. Ann Rheum Dis 1979;38(560):1. [Medline: 317239]
35. Ostelo RWJG, de Vet HC, Knol DL, van den Brandt PA. 24-item Roland-Morris Disability Questionnaire was preferred out of six functional status questionnaires for post-lumbar disc surgery. J Clin Epidemiol 2004 Mar;57(3):268-276. [doi: 10.1016/j.jclinepi.2003.09.005] [Medline: 15066687]
36. Dunn KM, Cherkin DC. The Roland-Morris Disability Questionnaire. Spine (Phila Pa 1976) 2007 Jan 15;32(2):287. [doi: 10.1097/01.brs.0000249551.00481.3d] [Medline: 17224833]
37. Roland M, Morris R. A study of the natural history of low-back pain. Part II: development of guidelines for trials of treatment in primary care. Spine (Phila Pa 1976) 1983 Mar;8(2):145-150. [Medline: 6222487]

38. Beurskens AJ, de Vet HC, Köke AJ. Responsiveness of functional status in low back pain: a comparison of different instruments. Pain 1996 Apr;65(1):71-76. [Medline: 8826492]
39. Roland-Morris Disability Questionnaire. 1983. URL: http://www.rmdq.org/ [accessed 2018-01-12] [WebCite Cache ID 6wPzCvX6M]
40. Downie WW, Leatham PA, Rhind VM, Wright V, Branco JA, Anderson JA. Studies with pain rating scales. Ann Rheum Dis 1978 Aug;37(4):378-381 [FREE Full text] [Medline: 686873]
41. Childs J, Piva S, Fritz J. Responsiveness of the numeric pain rating scale in patients with low back pain. Spine 2005;30(11):1. [Medline: 15928561]
42. Farrar JT, Young JP, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 2001 Nov;94(2):149-158. [Medline: 11690728]
43. de Vet HC, Ostelo R, Terwee C, van der Roer N, Knol D, Beckerman H, et al. Minimally important change determined by a visual method integrating an anchor-based and a distribution-based approach. Qual Life Res 2007;16:131-142. [Medline: 17033901]
44. Lauridsen HH, Hartvigsen J, Korsholm L, Grunnet-Nilsson N, Manniche C. Choice of external criteria in back pain research: does it matter? Recommendations based on analysis of responsiveness. Pain 2007 Sep;131(1-2):112-120. [doi: 10.1016/j.pain.2006.12.023] [Medline: 17276006]
45. Davidson M, Keating JL. A comparison of five low back disability questionnaires: reliability and responsiveness. Phys Ther 2002 Jan;82(1):8-24. [Medline: 11784274]
46. Deyo RA, Centor RM. Assessing the responsiveness of functional scales to clinical change: an analogy to diagnostic test performance. J Chronic Dis 1986;39(11):897-906. [Medline: 2947907]
47. Grotle M, Brox JI, Vøllestad NK. Concurrent comparison of responsiveness in pain and functional status measurements used for patients with low back pain. Spine (Phila Pa 1976) 2004 Nov 01;29(21):E492-E501. [Medline: 15507789]
48. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007 Jan;60(1):34-42. [doi: 10.1016/j.jclinepi.2006.03.012] [Medline: 17161752]
49. Jaeschke R, Singer J, Guyatt GH. Measurement of health status: ascertaining the minimal clinically important difference. Control Clin Trials 1989 Dec;10(4):407-415. [Medline: 2691207]
50. Froud R, Abel G. Using ROC curves to choose minimally important change thresholds when sensitivity and specificity are valued equally: the forgotten lesson of pythagoras. theoretical considerations and an example application of change in health status. PLoS One 2014;9(12):e114468 [FREE Full text] [doi: 10.1371/journal.pone.0114468] [Medline: 25474472]
51. ROCMIC: Stata module to estimate minimally important change (MIC) thresholds for continuous clinical outcome measures using ROC curves database on the Internet. URL: https://ideas.repec.org/c/boc/bocode/s457052.html [accessed 2018-08-29] [WebCite Cache ID 722gkYSm4]
52. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull 1979 Mar;86(2):420-428. [Medline: 18839484]
53. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. J Clin Epidemiol 2006 Oct;59(10):1033-1039. [doi: 10.1016/j.jclinepi.2005.10.015] [Medline: 16980142]
54. Sitgreaves R. Review of Intraclass correlation and the analysis of variance by E. A. Haggard. J Am Stat Assoc 1960;55:384-385.
55. Fleiss JL. The Design and Analysis of Clinical Experiments. New York: Wiley; 1986.
56. Stratford PW, Riddle DL. A Roland Morris Disability Questionnaire target value to distinguish between functional and dysfunctional states in people with low back pain. Physiother Can 2016;68(1):29-35 [FREE Full text] [doi: 10.3138/ptc.2014-85] [Medline: 27504045]
57. Ostelo RWJG, de Vet HC. Clinically important outcomes in low back pain. Best Pract Res Clin Rheumatol 2005 Aug;19(4):593-607. [doi: 10.1016/j.berh.2005.03.003] [Medline: 15949778]
58. Guyatt GH, Norman GR, Juniper EF, Griffith LE. A critical look at transition ratings. J Clin Epidemiol 2002 Sep;55(9):900-908. [Medline: 12393078]
59. Tilford J, Roberson P, Fiser D. Using lfit and lroc to evaluate mortality prediction models. Stata Technical Bulletin 1995;5:77-81.
60. Altman DG. Practical Statistics for Medical Research. London: Chapman & Hall; 1991.
61. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982 Apr;143(1):29-36. [doi: 10.1148/radiology.143.1.7063747] [Medline: 7063747]
62. Efron B, Tibshirani R. Bootstrap measures for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical Science (1) 1986:54-77.
63. Froud R. Improving Interpretation of Patient-Reported Outcomes in Low Back Pain Trials. London: Queen Mary University of London; 2010.
64. Gould W, Pitblado J. ACCUM: Stata module. 2001. URL: https://www.stata.com/support/faqs/statistics/bootstrapped-samples-guidelines/ [accessed 2018-01-12] [WebCite Cache ID 6wPzQpaXc]

65. Fawkes CA, Leach CMJ, Mathias S, Moore AP. Development of a data collection tool to profile osteopathic practice: use of a nominal group technique to enhance clinician involvement. Man Ther 2014 Apr;19(2):119-124. [doi: 10.1016/j.math.2013.08.006] [Medline: 24119310]
66. Fawkes CA, Leach CMJ, Mathias S, Moore AP. A profile of osteopathic care in private practices in the United Kingdom: a national pilot using standardised data collection. Man Ther 2014 Apr;19(2):125-130. [doi: 10.1016/j.math.2013.09.001] [Medline: 24139392]
67. Maughan EF, Lewis JS. Outcome measures in chronic low back pain. Eur Spine J 2010 Sep;19(9):1484-1494 [FREE Full text] [doi: 10.1007/s00586-010-1353-6] [Medline: 20397032]
68. Stratford PW, Binkley JM, Riddle DL, Guyatt GH. Sensitivity to change of the Roland-Morris Back Pain Questionnaire: part 1. Phys Ther 1998 Nov;78(11):1186-1196. [Medline: 9806623]
69. Beurskens AJ, de Vet HC, Köke AJ, van der Heijden GJ, Knipschild PG. Measuring the functional status of patients with low back pain: assessment of the quality of four disease-specific questionnaires. Spine (Phila Pa 1976) 1995 May 01;20(9):1017-1028. [Medline: 7631231]
70. Frost H, Lamb SE, Stewart-Brown S. Responsiveness of a patient specific outcome measure compared with the Oswestry Disability Index v2.1 and Roland and Morris Disability Questionnaire for patients with subacute and chronic low back pain. Spine (Phila Pa 1976) 2008 Oct 15;33(22):2450-2457. [doi: 10.1097/BRS.0b013e31818916fd] [Medline: 18824951]
71. Stratford PW, Binkley J, Solomon P, Gill C, Finch E. Assessing change over time in patients with low back pain. Phys Ther 1994 Jun;74(6):528-533. [Medline: 8197239]
72. Mannion AF, Junge A, Grob D, Dvorak J, Fairbank JCT. Development of a German version of the Oswestry Disability Index. Part 2: sensitivity to change after spinal surgery. Eur Spine J 2006 Jan;15(1):66-73 [FREE Full text] [doi: 10.1007/s00586-004-0816-z] [Medline: 15856340]
73. Coelho RA, Siqueira FB, Ferreira PH, Ferreira ML. Responsiveness of the Brazilian-Portuguese version of the Oswestry Disability Index in subjects with low back pain. Eur Spine J 2008 Aug;17(8):1101-1106 [FREE Full text] [doi: 10.1007/s00586-008-0690-1] [Medline: 18512083]
74. Brouwer S, Kuijer W, Dijkstra PU, Göeken LNH, Groothoff JW, Geertzen JHB. Reliability and stability of the Roland Morris Disability Questionnaire: intra class correlation and limits of agreement. Disabil Rehabil 2004 Feb 04;26(3):162-165. [doi: 10.1080/09638280310001639713] [Medline: 14754627]
75. Lauridsen HH, Hartvigsen J, Manniche C, Korsholm L, Grunnet-Nilsson N. Danish version of the Oswestry disability index for patients with low back pain. Part 2: Sensitivity, specificity and clinically significant improvement in two low back pain populations. Eur Spine J 2006 Nov;15(11):1717-1728. [doi: 10.1007/s00586-006-0128-6] [Medline: 16736202]
76. Salaffi F, Stancati A, Silvestri CA, Ciapetti A, Grassi W. Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. Eur J Pain 2004 Aug;8(4):283-291. [doi: 10.1016/j.ejpain.2003.09.004] [Medline: 15207508]
77. Janwantanakul P, Sihawong R, Sitthipornvorakul E, Paksaichol A. A screening tool for non-specific low back pain with disability in office workers: a 1-year prospective cohort study. BMC Musculoskelet Disord 2015 Oct 14;16:298 [FREE Full text] [doi: 10.1186/s12891-015-0768-y] [Medline: 26467434]
78. Scrimshaw SV, Maher C. Responsiveness of visual analogue and McGill pain scale measures. J Manipulative Physiol Ther 2001 Oct;24(8):501-504. [doi: 10.1067/mmt.2001.118208] [Medline: 11677548]
79. Parker SL, Adogwa O, Paul AR, Anderson WN, Aaronson O, Cheng JS, et al. Utility of minimum clinically important difference in assessing pain, disability, and health state after transforaminal lumbar interbody fusion for degenerative lumbar spondylolisthesis. J Neurosurg Spine 2011 May;14(5):598-604. [doi: 10.3171/2010.12.SPINE10472] [Medline: 21332281]
80. Lauridsen HH, Hartvigsen J, Manniche C, Korsholm L, Grunnet-Nilsson N. Responsiveness and minimal clinically important difference for pain and disability instruments in low back pain patients. BMC Musculoskelet Disord 2006 Oct 25;7:82 [FREE Full text] [doi: 10.1186/1471-2474-7-82] [Medline: 17064410]
81. Kovacs FM, Abraira V, Royuela A, Corcoll J, Alegre L, Cano A, et al. Minimal clinically important change for pain intensity and disability in patients with nonspecific low back pain. Spine (Phila Pa 1976) 2007 Dec 01;32(25):2915-2920. [doi: 10.1097/BRS.0b013e31815b75ae] [Medline: 18246018]
82. Riddle DL, Stratford PW, Binkley JM. Sensitivity to change of the Roland-Morris Back Pain Questionnaire: part 2. Phys Ther 1998 Nov;78(11):1197-1207. [Medline: 9806624]
83. Jordan K, Dunn KM, Lewis M, Croft P. A minimal clinically important difference was derived for the Roland-Morris Disability Questionnaire for low back pain. J Clin Epidemiol 2006 Jan;59(1):45-52. [doi: 10.1016/j.jclinepi.2005.03.018] [Medline: 16360560]
84. van der Roen N, Ostelo RWJG, Bekkering GE, van Tulder MW, de Vet HC. Minimal clinically important change for pain intensity, functional status, and general health status in patients with nonspecific low back pain. Spine (Phila Pa 1976) 2006 Mar 01;31(5):578-582. [doi: 10.1097/01.brs.0000201293.57439.47] [Medline: 16508555]
85. Rose G. The Strategy of Preventive Medicine. Oxford: Oxford University Press; 1992.
86. de Vet HC, Beckerman H, Terwee CB, Terluin B, Bouter LM. Definition of clinical differences. J Rheumatol 2006 Feb;33(2):434-435. [Medline: 16465677]

87. Costa LOP, Maher CG, Latimer J. Self-report outcome measures for low back pain: searching for international cross-cultural adaptations. Spine (Phila Pa 1976) 2007 Apr 20;32(9):1028-1037. [doi: 10.1097/01.brs.0000261024.27926.0f] [Medline: 17450079]
88. Mannion AF, Elfering A, Staerkle R, Junge A, Grob D, Semmer NK, et al. Outcome assessment in low back pain: how low can you go? Eur Spine J 2005 Dec;14(10):1014-1026. [doi: 10.1007/s00586-005-0911-9] [Medline: 15937673]
89. Parker SL, Mendenhall SK, Shau DN, Adogwa O, Anderson WN, Devin CJ, et al. Minimum clinically important difference in pain, disability, and quality of life after neural decompression and fusion for same-level recurrent lumbar stenosis: understanding clinical versus statistical significance. J Neurosurg Spine 2012 May;16(5):471-478. [doi: 10.3171/2012.1.SPINE11842] [Medline: 22324801]
90. Bird M, Callisaya ML, Cannell J, Gibbons T, Smith ST, Ahuja KD. Accuracy, validity, and reliability of an electronic visual analog scale for pain on a touch screen tablet in healthy older adults: a clinical trial. Interact J Med Res 2016 Jan 14;5(1):e3 [FREE Full text] [doi: 10.2196/ijmr.4910] [Medline: 26769149]
91. Underwood MR, Parsons S, Eldridge SM, Spencer AE, Feder GS. Asking older people about fear of falling did not have a negative effect. J Clin Epidemiol 2006 Jun;59(6):629-634. [doi: 10.1016/j.jclinepi.2005.09.014] [Medline: 16713526]
92. Bijur PE, Silver W, Gallagher EJ. Reliability of the visual analog scale for measurement of acute pain. Acad Emerg Med 2001 Dec;8(12):1153-1157. [Medline: 11733293]
93. Gallagher EJ, Bijur PE, Latimer C, Silver W. Reliability and validity of a visual analog scale for acute abdominal pain in the ED. Am J Emerg Med 2002 Jul;20(4):287-290. [Medline: 12098173]
94. Bishop FL, Lewis G, Harris S, McKay N, Prentice P, Thiel H, et al. A within-subjects trial to test the equivalence of online and paper outcome measures: the Roland Morris disability questionnaire. BMC Musculoskelet Disord 2010 Jun 08;11:113 [FREE Full text] [doi: 10.1186/1471-2474-11-113] [Medline: 20529332]
95. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017 Apr 19;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]
96. Mokkink LB, Prinsen CAC, Bouter LM, de Vet HC, Terwee CB. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. Braz J Phys Ther 2016 Jan 19;20(2):105-113 [FREE Full text] [doi: 10.1590/bjpt-rbf.2014.0143] [Medline: 26786084]
97. Froud R, Ellard D, Patel S, Eldridge S, Underwood M. Primary outcome measure use in back pain trials may need radical reassessment. BMC Musculoskelet Disord 2015 Apr 14;16:88 [FREE Full text] [doi: 10.1186/s12891-015-0534-1] [Medline: 25887581]
98. Fawkes C, Carnes D, Froud R. Introducing electronic PROM data collection into clinical practice. Learning the lessons from a pilot study. Physiotherapy 2017;103:e111.
99. Fawkes C. The Development, Evaluation, and Initial Implementation Of A National Programme For The Use and Collation Of Patient Reported Outcome Measures (Proms) In Osteopathic Back Pain Services In The UK. London: Queen Mary University of London; 2017.
100. Deeny SR, Steventon A. Making sense of the shadows: priorities for creating a learning healthcare system based on routinely collected data. BMJ Qual Saf 2015 Aug;24(8):505-515 [FREE Full text] [doi: 10.1136/bmjqs-2015-004278] [Medline: 26065466]
101. Celi LA, Davidzon G, Johnson AE, Komorowski M, Marshall DC, Nair SS, et al. Bridging the health data divide. J Med Internet Res 2016 Dec 20;18(12):e325 [FREE Full text] [doi: 10.2196/jmir.6400] [Medline: 27998877]

Abbreviations

AUC: area under the curve
COSMIN: Consensus-Based Standards for the Selection of Health Measurement Instruments
eNRS: electronic numerical rating scale
ePRO: electronic patient-reported outcome measure
eRMDQ: electronic Roland Morris Disability Questionnaire
eVAS: electronic visual analog scale
ICC: intraclass correlation coefficient
MIC: minimally important change
MDC: minimal detectable change
MDC $_{95}$: minimal detectable change at the 95% level
NRS: numerical rating scale
PROM: patient-reported outcome measure
RMDQ: Roland Morris Disability Questionnaire
ROC: receiver operator characteristic
VAS: visual analog scale

Edited by G Eysenbach; submitted 15.01.18; peer-reviewed by M Alshehri, A Riis; comments to author 30.03.18; revised version received 18.05.18; accepted 18.06.18; published 24.10.18.

Please cite as:
Froud R, Fawkes C, Foss J, Underwood M, Carnes D
Responsiveness, Reliability, and Minimally Important and Minimal Detectable Changes of 3 Electronic Patient-Reported Outcome Measures for Low Back Pain: Validation Study
J Med Internet Res 2018;20(10):e272
URL: http://www.jmir.org/2018/10/e272/
doi:10.2196/jmir.9828
PMID:30355556

Original Paper

Effectiveness of the Malnutrition eLearning Course for Global Capacity Building in the Management of Malnutrition: Cross-Country Interrupted Time-Series Study

Sunhea Choi[1], BEng, MSc, PhD; Ho Ming Yuen[2], BSc, MSc; Reginald Annan[3], BSc, MSc, PhD; Trevor Pickup[2], BSc; Andy Pulman[2], HNC, HND, MA, PhD; Michele Monroy-Valle[4], ND, MEd, MSc; Nana Esi Linda Aduku[3], BSc, MPhil; Samuel Kyei-Boateng[3], BSc, MPhil; Carmen Isabel Velásquez Monzón[4], ND, MA; Carmen Elisa Portillo Sermeño[5], ND, PGDip; Andrew Penn[2], BSc, MSc; Ann Ashworth[6], BSc, PhD, Hon FRCPCH; Alan A Jackson[2], BA, MA, MB, BChir, MD, FRCP, FRCPCH, FRCPath, RNutr

[1]Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, United Kingdom
[2]Faculty of Medicine, University of Southampton, Southampton, United Kingdom
[3]Department of Biochemistry and Biotechnology, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana
[4]Facultad de Ciencias de la Salud, Universidad Rafael Landívar, Ciudad de Guatemala, Guatemala
[5]Department of Nutrition and Diets, Hospital Nacional "Dr Juan José Fernandez" Zacamil, San Salvador, El Salvador
[6]Department of Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom

Corresponding Author:
Sunhea Choi, BEng, MSc, PhD
Human Development and Health
Faculty of Medicine
University of Southampton
Mailpoint 820, Level B, South Academic Block
Southampton General Hospital
Southampton, SO16 6YD
United Kingdom
Phone: 44 23 8120 4330
Email: s.choi@southampton.ac.uk

Abstract

Background: Scaling up improved management of severe acute malnutrition has been identified as the nutrition intervention with the largest potential to reduce child mortality, but lack of operational capacity at all levels of the health system constrains scale-up. We therefore developed an interactive malnutrition eLearning course that is accessible at scale to build capacity of the health sector workforce to manage severely malnourished children according to the guidelines of the World Health Organization.

Objective: The aim of this study was to test whether the malnutrition eLearning course improves knowledge and skills of in-service and preservice health professionals in managing children with severe acute malnutrition and enables them to apply the gained knowledge and skills in patient care.

Methods: This 2-year prospective, longitudinal, cross-country, interrupted time-series study took place in Ghana, Guatemala, El Salvador, and Colombia between January 2015 and February 2017. A subset of 354 in-service health personnel from 12 hospitals and 2 Ministries of Health, 703 preservice trainees from 9 academic institutions, and 204 online users participated. Knowledge gained after training and retention over time was measured through pre- and postassessments comprising questions pertaining to screening, diagnosis, pathophysiology and treatment, and prevention of malnutrition. Comprehension, application, and integration of knowledge were tested. Changes in perception, confidence, and clinical practice were assessed through questionnaires and interviews.

Results: Before the course, awareness of the World Health Organization guidelines was 36.73% (389/1059) overall, and 26.3% (94/358) among in-service professionals. The mean score gain in knowledge after access to the course in 606 participants who had pre- and postassessment data was 11.8 (95% CI 10.8-12.9; *P*<.001)—a relative increase of 41.5%. The proportion of participants who achieved a score above the pass mark posttraining was 58.7% (356/606), compared with 18.2% (110/606) in pretraining. Of the in-service professionals, 85.9% (128/149) reported applying their knowledge by changing their clinical practice in screening, assessment, diagnosis, and management. This group demonstrated significantly increased retained knowledge 6 months after training (mean difference [SD] from preassessment of 12.1 [11.8]), retaining 65.8% (12.1/18.4) of gained knowledge from the training. Changes in the management of malnutrition were reported by trained participants, and institutional, operational, and policy changes were also found.

Conclusions: The malnutrition eLearning course improved knowledge, understanding, and skills of health professionals in the diagnosis and management of children with severe acute malnutrition, and changes in clinical practice and confidence were reported following the completion of the course.

(J Med Internet Res 2018;20(10):e10396) doi:10.2196/10396

KEYWORDS

eLearning; severe acute malnutrition; WHO guidelines for malnutrition; capacity building; staff development; quality improvement; nutrition training and education

Introduction

Background

Undernutrition (low weight-for-age) is associated with 3.1 million child deaths annually [1]. In the 2013 Lancet Series, scaling up the management of severe acute malnutrition (SAM, ie, severe wasting, severe wasting with edema, or edema) was identified as the nutrition intervention with the largest potential to reduce child mortality [2]. Lack of operational capacity at all levels of the health sector, however, constrains scale-up [3]. In countries most affected by SAM, training and curricula are outdated or nonexistent and misaligned with strategic and operational needs [4], leading to a workforce that is ill equipped to identify and treat malnourished children [5]. This limited knowledge and competency of health professionals in managing children with SAM led the International Pediatric Association to pass a resolution in 2010, stating that all pediatricians and related health professionals should have the identification and treatment of severe malnutrition as a core competency [6]. It was against this background that the International Malnutrition Task Force of the International Union of Nutritional Sciences joined with the University of Southampton to develop an eLearning training course on “Caring for infants and young children with severe malnutrition” that could be accessed at scale through the internet (Multimedia Appendix 1). The project goal was to build the capacity of the health sector workforce globally to manage SAM and reduce child deaths. The rapid spread of the internet and access to information technology (IT) across the developing world provides exciting new opportunities for delivering training in a way that was not possible before. If utilized effectively, we believe eLearning can make a significant contribution to building capacity for improved malnutrition management, supporting the aims of the Scaling Up Nutrition Movement, through a high-volume training of relevant personnel around the world.

The World Health Organization (WHO) guidelines for the management of SAM have been shown to be demonstrably effective when implemented appropriately [2,7-9], and well-trained, motivated staff have been shown to reduce SAM mortality in practice [9-11]. It has been suggested that by harnessing the potential of technology, particularly the internet, it should be possible to train health professionals at scale [12-14]. eLearning is at least as effective as traditional teaching in facilitating knowledge and skills learning [15-17], and it can be cost effective [18]. Learning is an active process of constructing knowledge (constructivist view) [19,20], and instruction is a process of supporting that construction [21]. Our course offers scenario-based eLearning, which aims to provide contextualized learning of patient care through malnutrition cases and their management. It creates a learning environment that promotes *deep* rather than *surface* learning [22-26], in which learners can actively construct new and meaningful knowledge [27-29]. Such an environment enables the learner to see the *relevance* of the new learning content in their context and situation and relate it to their existing knowledge or experience. The course incorporates aspects of problem-based learning [30] and situated learning with authentic tasks [31]. These approaches aim to facilitate active learning with contextually relevant tasks through which the learner can acquire, integrate, and apply knowledge and skills and subsequently be able to apply the practiced knowledge and skills in their clinical work. An example is virtual patients, which were created to facilitate contextualized learning of basic and clinical sciences for early medical training [32] and case-based learning through branching scenarios for clinical reasoning and decision-making skills training [33-37]. Adaptive learning aims to provide tailored training to suit the individual needs of the learner to maximize performance, and it is hypothesized that the greater the degree of adaptation and the inclusion of content adaptation, the greater is its effectiveness [38]. Factors such as learners’ knowledge levels, difficulty level of learning content, characteristics of the learner (learning styles, cognitive styles), and their preferences are used in adaptive eLearning systems to customize the content, navigation, presentation (media and layout), and materials/tools [38,39].

Despite widespread use, there are important gaps in knowledge regarding eLearning. For example, little is known about the effectiveness of eLearning on patient outcomes [16], and the environments in which eLearning occurs tend to be limited. For example, in a systematic review of undergraduate health professional education commissioned by the WHO, only 5 of 49 eLearning studies were from low- or middle-income countries (Brazil, China, and Thailand), and only 14 considered professions allied to medicine such as nursing [15]. The authors of the review recommended that future research should assess the outcomes of eLearning in health care training (1) in low-and middle-income countries and (2) among professionals allied to medicine. They also recommended evaluating the impact of eLearning on long-term retention of knowledge and skills. Our study has relevance with each of these knowledge gaps.

Study and Objectives

Since its launch, the malnutrition eLearning course has been used by over 14,000 health professionals, trainees, and educators, and positive anecdotal feedback has been received [40]. Knowledge scores embedded in each module show improvement on completion of the modules; however, although they are encouraging, they do not adequately show whether the knowledge gained is actually applied in practice and whether this leads to improved management of severely malnourished children. This evaluation was therefore undertaken to investigate whether (1) the malnutrition eLearning course improves the knowledge and skills in managing children with SAM of in-service and preservice health professionals (*think differently*) and enables them to apply the gained knowledge and skills in patient care (*act differently*) and (2) their application of the gained knowledge and skills leads to improved clinical practice and outcomes for severely malnourished children in resource-poor countries. In this paper, we address the first hypothesis. The second hypothesis is addressed in a separate paper.

Methods

Design

This is a prospective, longitudinal, cross-country, interrupted time-series study investigating the effectiveness of an educational intervention in facilitating the gain of knowledge and skills and changes in clinical practice, which could lead to improved care of children with SAM. Kirkpatrick training evaluation model [41] was applied to the study design. It comprises 4 levels: reaction, learning, behavior, and results.

Intervention

Course Developers

The eLearning course was developed by members of the International Malnutrition Task Force and the University of Southampton. The content was prepared mainly by RA and AA, and the interactive, task-based design was created by SC. The host platform and course were developed by SC and the University of Southampton Faculty of Medicine eLearning team. For many years, AA and AAJ served as advisers to the WHO on inpatient management of severe malnutrition and on training of health professionals. AA also served as a facilitator for WHO regional training courses to improve the inpatient management of severe malnutrition.

Development Process

The design of the course is scenario-based and interactive to actively support and engage the user in the learning process. The course uses a range of rich media; however, it can be run on a low specification computer with a limited internet speed. In addition, a stand-alone CD version of the course supports the areas where internet access is not possible. The content is set at a level suitable for the primary target groups, namely in-service and preservice health professionals who are or will be working with undernourished children. Secondary target groups are educators and trainers in medical, nursing, and health science schools and organizations with responsibility for preservice training. The target regions are those where the prevalence of malnutrition is high, notably sub-Saharan Africa, India and the subcontinent, and Latin America [42]. Upon the completion of the course development in 2010, a field test (beta testing) was conducted in Uganda [43] (Multimedia Appendix 2) to assess the effectiveness of the course and the appropriateness of its delivery for the target users and regions. Overall, 86 participants, including doctors, medical students, nurses, midwives, and nutritionists, participated in the study, and the results indicated that the course design supports learning for different health professionals, and the course is accessible with limited internet bandwidth.

Revisions and Updates

After the field test in Uganda, adjustments were made to eliminate ambiguities and ensure smooth functioning of the course. In 2014, the content was slightly revised to match the updated WHO case-management guidelines. No further changes were made either before or during the study. Currently, the course is being reimplemented from Adobe Flash to html5, and the new version will be made freely available in October 2018.

Quality Assurance

The course content conforms to the WHO case-management guidelines [44-46]. Standard anthropometric techniques are used for assessing nutritional status. The accuracy of the content was checked by 3 experienced public health nutritionists working independently.

Replicability

The course is freely available from the University of Southampton nutrition portal [47]. The design solution has been disseminated through presentations, poster, and handouts at eLearning events (exhibitions and conferences) held by the eLearning team in 2009 and 2011. The dissemination materials are available from the medicine eLearning website [48]. The technical solutions designed for the nutrition portal and course implementation are available in published papers [49,50].

Access

There are no preconditions and anyone can access the course. Study participants gained access by registering for the course in the University of Southampton nutrition portal, either via their own laptop or from 1 of their institution's computers, or by using a CD version.

Mode of Delivery and Content

The course offers interactive learning in 3 modules: Module 1: Definition and classification of malnutrition; Module 2: How to identify children with malnutrition; and Module 3: How to manage children with malnutrition. The course facilitates learning on how to differentiate between chronic and acute malnutrition; pathophysiological changes in malnutrition and implications for treatment; how to assess and screen children for malnutrition and interpret the results for action; how to manage malnourished children using the WHO Ten Steps;

hospital compared with community-based management; the importance of an integrated approach between hospital and community; and how to support mothers and caregivers to prevent the recurrence of malnutrition.

The course is designed based on a constructivist view of learning—"an active process of constructing knowledge" [19-21]—and aims to facilitate the activities required for learning to occur, which are apprehending structure, integration, application, and reflection [29,51-53]. The 2 key overarching design strategies are (1) scenario- and task-based learning activities using authentic cases to support an interactive process of acquiring, linking, integration, and application of knowledge and skills and (2) activities embedded within the content, that is, revisable reflective questions and conversational style to promote active reflection and cognitive dialogue. Figure 1 shows illustrative snapshots of the 2 design strategies.

Each module usually takes 2 to 3 hours to complete but is asynchronous as users take the course in their own time. Each module comprises several interlinked subunits, and users can track their progress through a set of multiple-choice questions at the beginning and end of each module. Users can navigate within the course as desired.

Human Involvement and Prompts

No support or assistance was given to participants while they were taking the course. The participants were introduced to the course at the end of the prestudy data collection and asked to complete the course within 3 weeks. There were no cointerventions or prompts.

Participants and Setting

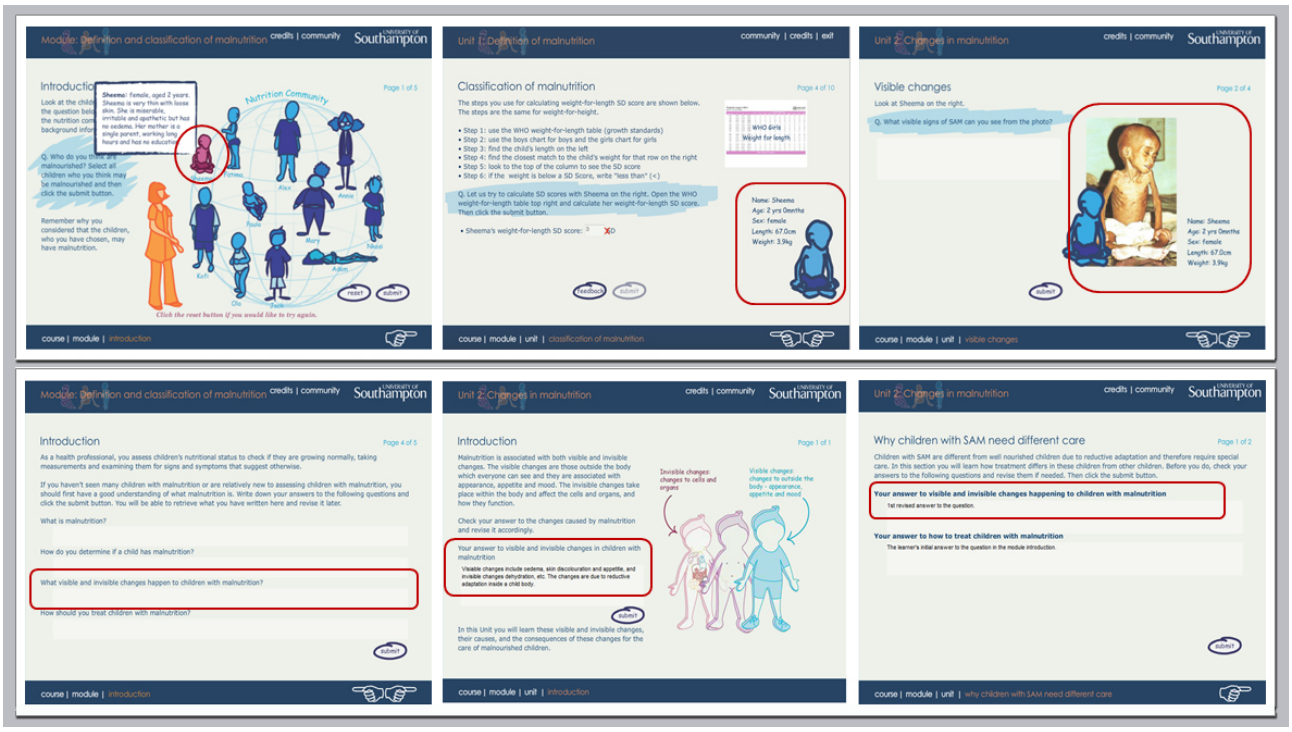
A subset of target course users and geographic locations was selected for the study that was conducted face-to-face in Ghana and in Guatemala, El Salvador, and Colombia (grouped for convenience as Latin America) and also globally online, between January 2015 and February 2017. Study participation was voluntary. The study consisted of 2 groups: (1) in-service health professionals and preservice trainees studying health science subjects in Ghana and Latin America (center-based group) and (2) a self-selected group of global users who elected to take the course during a defined period of time (remote learning group).

Center-Based Group

Health care and academic institutions in Ghana (Ashanti region) and Latin America (El Salvador, Guatemala, and Colombia) were invited to participate. Health care institutions providing pediatric care (hospitals and community health centers) and academic institutions offering health science training programs (universities and training colleges) were eligible to participate.

In Ghana, 10 hospitals and their linked community health centers, and 8 academic institutions were contacted. Of these, 9 hospitals (and linked health centers) and 7 academic institutions accepted the invitation and participated in the evaluation. In Latin America, 3 of the 4 contacted hospitals, 2 of the 8 academic institutions, and the Ministries of Health for Guatemala and El Salvador accepted the invitation. Educators at the participating academic institutions were invited to 1 of the 3 2-day workshops during which they were introduced to the malnutrition eLearning course, and they developed action plans to initiate the course at their home institutions. All participants in Latin America and preservice participants in Ghana were introduced to the course by the research team or trained educators at their institutions.

Figure 1. Overarching design strategies: illustrative snapshots of scenario and task-based activities (top row) and revisable reflective questions (bottom row). The top row shows the introduction of malnutrition concepts and classification leading to practical application through Sheema, and the bottom row shows a reflective question, "What visible and invisible changes happen to children with malnutrition?", appearing at 3 different learning points.



Look at the child… the question bel… the nutrition com… background infor…

Sheema: female, aged 2 years. Sheema is very thin with loose skin. She is miserable, irritable and apathetic but has no oedema. Her mother is a single parent, working long hours and has no educatio

Q. Who do you think are malnourished? Select all children who you think may be malnourished and then click the submit button.

Remember why you considered that the children, who you have chosen, may have malnutrition.

Click the reset button if you would like to try again.

The steps you use for calculating weight-for-length SD score are shown below. The steps are the same for weight-for-height.

- Step 1: use the WHO weight-for-length table (growth standards)
- Step 2: use the boys chart for boys and the girls chart for girls
- Step 3: find the child's length on the left
- Step 4: find the closest match to the child's weight for that row on the right
- Step 5: look to the top of the column to see the SD score
- Step 6: if the weight is below a SD Score, write "less than" (<)

Q. Let us try to calculate SD scores with Sheema on the right. Open the WHO weight-for-length table top right and calculate her weight-for-length SD score. Then click the submit button.

- Sheema's weight-for-length SD score: 3 ✗SD

Name: Sheema
Age: 2 yrs 0mnths
Sex: female
Length: 67.0cm
Weight: 3.9kg

Look at Sheema on the right.

Q. What visible signs of SAM can you see from the photo?

Name: Sheema
Age: 2 yrs 0mnths
Sex: female
Length: 67.0cm
Weight: 3.9kg

As a health professional, you assess children's nutritional status to check if they are growing normally, taking measurements and examining them for signs and symptoms that suggest otherwise.

If you haven't seen many children with malnutrition or are relatively new to assessing children with malnutrition, you should first have a good understanding of what malnutrition is. Write down your answers to the following questions and click the submit button. You will be able to retrieve what you have written here and revise it later.

What is malnutrition?

How do you determine if a child has malnutrition?

What visible and invisible changes happen to children with malnutrition?

How should you treat children with malnutrition?

Malnutrition is associated with both visible and invisible changes. The visible changes are those outside the body which everyone can see and they are associated with appearance, appetite and mood. The invisible changes take place within the body and affect the cells and organs, and how they function.

Check your answer to the changes caused by malnutrition and revise it accordingly.

Your answer to visible and invisible changes in children with malnutrition

Visiable changes include oedema, skin discolouration and appetite, and invisible changes dehydration, etc. The changes are due to reductive adaptation inside a child body.

In this Unit you will learn these visible and invisible changes, their causes, and the consequences of these changes for the care of malnourished children.

Children with SAM are different from well nourished children due to reductive adaptation and therefore require special care. In this section you will learn how treatment differs in these children from other children. Before you do, check your answers to the following questions and revise them if needed. Then click the submit button.

Your answer to visible and invisible changes happening to children with malnutrition

1st revised answer to the question.

Your answer to how to treat children with malnutrition

The learner's initial answer to the question in the module introduction.

They were then given 3 weeks to complete the course either online or using the CD version installed on their institutional computers. For the Ghana in-service participants, a 2-day, self-directed training session was organized at each participating hospital and facilitated by the research team.

Remote Learning Group

Between October 2015 and January 2016, new users of the malnutrition eLearning course were invited at the time of enrollment to take part in the study. The inclusion criteria were that they were (1) in-service health professionals who worked closely with children or (2) preservice health professionals studying medicine or allied health science subjects and were not part of the participating institutions in Ghana and Latin America. Of the 322 who responded, 263 met the inclusion criteria. They came from 38 health care, academic, and nongovernment organizations across 40 countries.

Baseline (pre) data were collected before the modules were taken and follow-up data were captured at 3 time points over the period of 1 year: post (immediately after), 6 months, and 12 months after the training.

Ethical Approval

This study was reviewed and approved by the ethics committees of the University of Southampton, United Kingdom (Ethics ID:12872), Komfo Anokye Teaching Hospital and Kwame Nkrumah University of Science and Technology, Ghana, and the Universidad Rafael Landívar, Guatemala. Informed consent was obtained from participants and institutions before the commencement of the study.

Data Collection

The effectiveness of the malnutrition eLearning course for learning was evaluated using a mixture of quantitative and qualitative methods comprising assessments, questionnaires, and interviews and focus groups with individual participants.

Assessments (Pre, Post, and 6 Months)

To measure the participants' gain in knowledge from the course immediately after the training and retention of the gained knowledge 6 months later, 2 sets of comparable assessments were prepared for the baseline and follow-up phases. One assessment was assigned for the prestudy and the other for the post and 6-month follow-up studies. Each consisted of 32 questions on key topics identified from the course, and the questions were prepared to test comprehension, application, and integration of knowledge. Participants ranged from medical doctors to community health workers, and differences in existing knowledge between professional groups as well as within each group were anticipated. A standard setting procedure using the Ebel method [54], which allows an expected standard to be set in the form of a pass mark based on the difficulty and importance of each question, was carried out for each assessment. The pass marks established for the pre- and follow-up assessments were 37.4% and 36.4%, respectively.

For the center-based group, assessments were conducted *exam-style* within a set time under supervision by members of the research team. The total score for each assessment was calculated based on each question carrying the same weight. For the remote learning group, equivalent online versions of the assessments were prepared. To minimize any carryover effect between assessments, scores were released to participants only after the final 6-month assessment.

The assessment scores were analyzed to determine (1) participants' acquisition of core knowledge and skills from the training, (2) how many and who acquired the pass marks, and (3) retained knowledge at 6-month follow-up. Key influencing factors considered were profession, prior training in the management of SAM, adherence to the WHO guidelines at baseline, partial or full completion of the malnutrition eLearning course, and application of knowledge in the follow-up period.

Questionnaires (Pre, Post, 6 Months, and 12 Months)

Questionnaires, consisting of closed and open-ended questions, were administered at 4 time points (pre, post, 6 months, and 12 months) with a combination of similar (to investigate changes over time) and time-specific questions. These questionnaires explored whether participants completed the course, perceived changes in knowledge and understanding of malnutrition and its management, application of knowledge and associated outcomes, confidence in performing relevant tasks, changes in clinical practice made by in-service participants and resulting policy changes at their health care facilities, and benefits gained from the training. These questionnaires were administered face-to-face for the center-based group and online for the remote learning group.

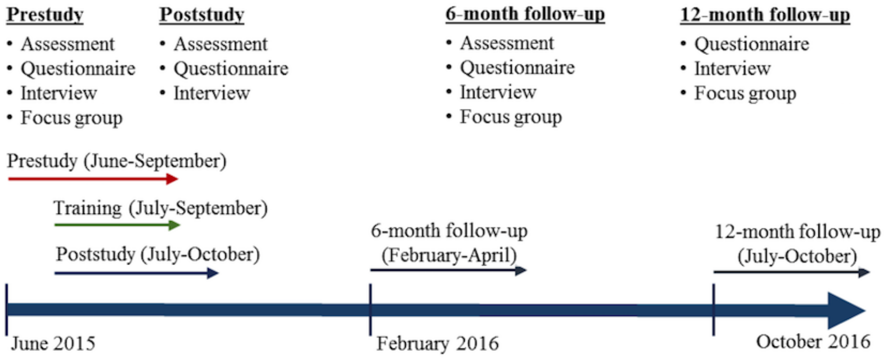
Interviews (Pre, Post, 6 Months, and 12 Months) and Focus Groups (Pre and 12 Months)

Semistructured interviews were conducted with the participants who volunteered to share further details of their experience, namely what they gained from the course, if and where they applied their gained knowledge and associated outcomes, if and how their perceptions about malnutrition and its management changed, and changes in the management of SAM made by them and observed at their workplaces. The interviews were audio-recorded and transcribed verbatim. Focus groups were conducted in Latin America with similar aims and structure to the interviews. Each consisted of 3 to 6 participants and was audio-recorded and transcribed verbatim. Figure 2 summarizes the data collection methods used.

Relation Between Knowledge Gain, Application, and Confidence in Patient Care

Improvement in knowledge and understanding was reported in 9 topic areas, from “types and classification of malnutrition” to “inpatient therapeutic care,” and the scores were aggregated and averaged to represent the overall response. Improvement in confidence in performing tasks was reported in 6 areas, from “screening children for malnutrition” to “managing children with severe acute malnutrition using WHO Ten Steps,” and the scores were aggregated and averaged to represent the overall response. Participants' gain in knowledge, its application in patient care, and reported increase in confidence from the training and subsequent application of the gained knowledge were assessed over time.

Figure 2. Study timeline and data collection methods used.



Our hypothesis was that the relation between improvement in knowledge and confidence would be weak at postassessment but would improve positively over time through the application of knowledge and experience of positive outcomes. The trend and correlation were explored at post, 6-month, and 12-month follow-up study points.

Data Management

Data management protocols (data collection, cleaning, and principles of analysis), research guides, and database templates were prepared at the start of the study and made available to the United Kingdom, Ghana, and Latin America teams through SharePoint. Secure research data storage was provided by the University of Southampton. Data were collected face-to-face and online by the 3 teams. The dataset collected from each team went through an independent cleaning process, and they were then merged into a combined dataset.

Statistical Analysis

Statistical analysis was performed on the quantitative data from the assessments and questionnaires. Assessment scores (out of 32) were converted into percentages. Questions from the questionnaires where participants needed to rate a series of statements using a Likert scale (1 to 5) regarding their current or improved level of knowledge, understanding, and confidence toward the malnutrition topics were aggregated and averaged to represent their overall response in the corresponding area. Cronbach alpha was used to measure their internal consistency, and the values ranged from .88 to .92.

Summary statistics were presented based on the types of variables. Continuous variables and the differences between paired continuous variables were assessed for normality using histograms. Paired sample *t* tests were performed when comparing the pre-, post-, and 6-month assessment scores. Complete-case analysis was also performed when assessing the changes over time across all 3 assessments. Subgroup analyses were performed by country, profession, and institution type. Spearman rank correlation was used to assess the relation between knowledge gain and improved confidence. Observations with missing data in the relevant variables under investigation were automatically discarded during the above analyses. Statistical significance was set at 5%. SPSS Statistics for Windows, Version 24.0 (IBM Corp, Armonk, NY) was used to perform these analyses.

Qualitative Analysis

Qualitative data from open-ended items in the questionnaires, interviews, and focus groups were analyzed using thematic analysis [55]. Transcripts were read and reread to aid familiarity with the data, and then data were coded to identify noteworthy findings. Codes were collated into categories to represent the dataset as a whole. Cohen kappa [56] was used to measure reproducibility across different coders (SC and APu).

Results

Participants

Of 1261 participants, 72.56% (915/1261) were from institutions in Ghana, 11.26% (142/1261) from Latin America, and 16.18% (204/1261) were remote online users. Table 1 shows the number of participants at each data collection point. Individual participants took part in 1 or more data collection activities.

Table 2 shows the characteristics of the participants, of whom 796 were preservice and 465 were in-service professionals. Of the in-service professionals, the majority (86.3%, 340/394) had regular close involvement with SAM children, but fewer than half of these (38.7%, 127/328) said they had received SAM training in the past. Awareness of the WHO Ten Steps was low (36.73%, 389/1059), especially among in-service professionals with only 26.3% (94/358) being aware of the guidelines. Only 20.7% (40/193) in-service professionals had received training about the WHO Ten Steps.

Access to a computer or laptop at home or at work was 100% (142/142) and 77.6% (672/866), respectively, among participants in Latin America and Ghana.

Knowledge Gain From Training and Retention at 6 Months

Table 3 shows the gain in knowledge posttraining. Of 1261 participants, 606 took both the pre- and postassessments. The gain in knowledge was compared in relation with country, profession, and extent to which participants had completed the course. The overall mean score gain in knowledge was 11.8 (*P*<.001)—a relative increase of 41.5%. Considering only those who completed the course, the mean gain was 14 points with a relative increase of 47.8%.

Table 1. Number of participants at each data collection point.

Methods[a]	Number of participants			
	Prestudy[b]	Poststudy[c]	6-month follow-up[d]	12-month follow-up[e]
Assessment				
Ghana	864	539	464	—
Latin America	141	60	109	—
Remote learning group	181	35	—	—
Questionnaire				
Ghana	895	548	447	249
Latin America	142	88	101	100
Remote learning group	100	16	—	5
Interview				
Ghana	33	14	22	19
Latin America	4	0	14	4
Remote learning group	4	1	—	0
Focus group				
Latin America	59[f]	—	—	14[g]

[a] Participant numbers overlap between data collection methods.

[b] June to September 2015.

[c] July to October 2015.

[d] February to April 2016.

[e] July to October 2016.

[f] 13 groups.

[g] 7 groups.

Of those who reported prior training in SAM management (Table 2), 28.1% (101/359) achieved the preassessment pass mark compared with 13.2% (89/672) who reported no prior training. Of those who reported following the WHO guidelines in their work, 57% (26/46) achieved the preassessment pass mark. Overall, the proportion of pre- and postassessment participants who achieved scores above the pass mark posttraining was 58.7% (356/606), compared with 18.2% (110/606) pretraining. Close to two-thirds (65.5%, 271/414) of the participants who completed the course obtained the pass mark.

Retention of gained knowledge at 6 months is shown in Figure 3 for the 332 participants who participated in all 3 assessments. Although overall there was some loss of knowledge by 6 months, the participants' knowledge remained significantly higher than pretraining (mean difference=7.1, 95% CI 5.9-8.4; *P*<.001) and the loss between posttraining and 6 months varied depending on whether or not participants applied their knowledge in clinical practice. For example, in-service participants who applied their knowledge retained 66% of the gained knowledge (mean scores for pre, post, and 6 months are 28.5, 46.8, and 40.5, respectively; retained knowledge=12.1/18.4) compared with 39% (mean scores for pre, post, and 6 months are 29.3, 42.2, and 34.3, respectively; retained knowledge=5.0/12.9) among those who did not apply their knowledge.

Knowledge Application and Changes in Clinical Practice

Participants were asked to report whether they had applied the knowledge, and 85.9% of in-service professionals who took the course (128/149) reported to have applied the gained knowledge in practice. At 6 months and 12 months, 256 (51.8%, 256/494) and 143 (68.8%, 143/208) participants reported to have applied knowledge in practice and provided 528 and 366 accounts of where they applied it, respectively (up to 3 accounts per respondent). All nutritionists reported applying knowledge in their practice, with the next highest group being nurses (88%, 37/42). Table 4 presents the rankings of the main areas of application. The rankings were similar at the 2 periods, and the most common applications were related to identification and treatment of SAM.

Detailed information around changes in clinical practice was sought at 12 months from in-service and graduate preservice participants. Of 130 respondents, 115 (88.5%, 115/130) stated that they had changed their practice in line with the WHO guidelines. The reported changes are summarized in Multimedia Appendix 3.

Table 2. Demographics of individual participants.

Variable[a]	Ghana (N=915), n (%)	Latin America (N=142), n (%)	Remote learning group (N=204), n (%)	Total (N=1261), n (%)
Profession				
Preservice (student)	597 (65.2)	106 (74.6)	93 (45.6)	796 (63.12)
Medical doctor	4 (0.4)	28 (19.7)	12 (5.9)	44 (3.49)
Nurse and midwife	228 (24.9)	—	8 (3.9)	236 (18.72)
Nutritionist	21 (2.3)	8 (5.6)	16 (7.8)	45 (3.57)
Public health	25 (2.7)	—	12 (5.9)	37 (2.93)
Other	40 (4.4)	—	63 (30.9)	103 (8.17)
Health care and academic institutions				
Hospital-based	186 (20.3)	28 (19.7)	—	214 (16.97)
Community-based	132 (14.4)	8 (5.6)	—	140 (11.10)
Universities	213 (23.3)	106 (74.6)	—	319 (25.30)
Training colleges	384 (42.0)	—	—	384 (30.45)
Remote learning group	—	—	204 (100)	204 (16.18)
Regular close involvement with SAM[b] children (in-service only)				
Yes	265 (86.9)	34 (94.4)	41 (77.4)	340 (86.3)
No	40 (13.1)	2 (5.6)	12 (22.6)	54 (13.7)
Aware of the WHO[c] Ten Steps for management of SAM				
Yes	287 (35.0)	50 (35.5)	52 (53.6)	389 (36.73)
No	534 (65.0)	91 (64.5)	45 (46.4)	670 (63.27)
Received training in the management of SAM in the past				
Yes	294 (34.2)	61 (43.0)	30 (30.6)	385 (35.03)
No	565 (65.8)	81 (57.0)	68 (69.4)	714 (64.97)
Received training about the WHO Ten Steps				
Yes	107 (24.2)	32 (53.3)	14 (30.4)	153 (27.9)
No	335 (75.8)	28 (46.7)	32 (69.6)	395 (72.1)
Following the WHO Ten Steps at work (in-service only)				
Yes	33 (28.2)	6 (35.3)	15 (71.4)	54 (34.8)
No	84 (71.8)	11 (64.7)	6 (28.6)	101 (65.2)

[a] Totals do not always add up to the number of participants as some questions were not answered by all.

[b] SAM: severe acute malnutrition.

[c] WHO: World Health Organization.

Table 3. Gain in knowledge post- versus preassessments.

Variable	Total participants (N)	Pre, mean (SD)	Post, mean (SD)	Post-pre difference	
				Mean (95% CI)	*P* value
Overall	606	28.4 (10.7)	40.2 (13.7)	11.8 (10.8 to 12.9)	<.001
Country					
Ghana	512	27.0 (9.9)	39.8 (13.4)	12.8 (11.7 to 13.9)	<.001
Latin America	60	32.6 (8.0)	42.4 (12.2)	9.7 (6.5 to 13.0)	<.001
Remote learning group	34	41.1 (15.1)	42.6 (19.5)	1.5 (−4.7 to 7.6)	.63
Total participants	606	28.4 (10.7)	40.2 (13.7)	11.8 (10.8 to 12.9)	<.001
Profession					
Preservice (student)	316	28.7 (10.1)	37.4 (14.1)	8.7 (7.2 to 10.2)	<.001
Medical doctor	6	44.3 (16.4)	52.1 (15.1)	7.8 (−2.8 to 18.4)	.12
Nurse and midwife	200	24.8 (8.2)	40.8 (10.6)	16.0 (14.5 to 17.6)	<.001
Nutritionist	21	41.1 (11.7)	55.4 (13.7)	14.3 (7.6 to 21.0)	<.001
Public health	22	31.7 (8.9)	45.5 (14.3)	13.8 (7.9 to 19.6)	<.001
Other	41	33.0 (14.8)	46.5 (14.5)	13.5 (9.5 to 17.5)	<.001
Total participants	606	28.4 (10.7)	40.2 (13.7)	11.8 (10.8 to 12.9)	<.001
Malnutrition eLearning use during the training					
Complete	397	29.1 (10.1)	43.0 (12.5)	14.0 (12.7 to 15.2)	<.001
In progress	86	27.8 (11.5)	36.6 (15.0)	8.8 (5.9 to 11.6)	<.001
Incomplete and stopped	82	22.7 (7.3)	32.1 (11.3)	9.3 (6.7 to 12.0)	<.001
Total participants	565[a]	27.9 (10.2)	40.4 (13.5)	12.5 (11.4 to 13.6)	<.001

[a] A total of 41 participants did not respond to this question during the postintervention stage when asked if they had used the malnutrition eLearning during the training.

Figure 3. Mean (SD) assessment scores pre, post, and 6 months for the 332 participants who had all 3 assessments.

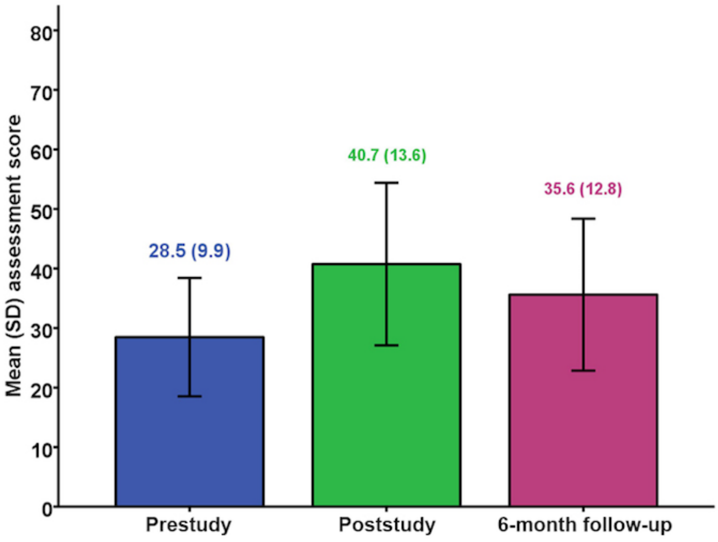


Table 4. Rankings of knowledge application reported by participants at 6 months and 12 months.

Category	Description	6 months (N=256)[a]	12 months (N=143)[a]
Identification of severe malnutrition	Using indicators (midupper arm circumference, weight-for-height) and clinical signs to assess and classify malnutrition	1	1
Treatment and management	Following the WHO[b] Ten Steps to treat children for severe acute malnutrition	2	2
Examining for clinical signs	Examining children for signs of malnutrition and associated conditions	3	4
Taking measurements	Taking weight, height, length, midupper arm circumference correctly	4	3
Screening for malnutrition	Screening as part of growth monitoring, home visits and outreach programs, and at outpatient department and wards.	5	7
Counseling mothers	Counseling about hygiene, feeding, causes and signs of malnutrition, and how to prevent malnutrition.	6	5
Admission criteria and management option	Applying WHO criteria for admission and deciding whether admission should be to inpatient or community-based care	7	8
Other	Training other health professionals, promoting WHO guidelines, supervision, and teaching family and friends	8	6

[a] Respondents were asked to state important ways (up to 3) in which they had applied their new knowledge. Only 1 account/category was counted for each respondent. Of the 528 accounts, 449 were counted at 6 months and 327 of 366 accounts were counted at 12 months.

[b] WHO: World Health Organization.

Table 5. Policy and operational changes reported by in-service participants in Ghana.

Variable	In-service[a] Hospital-based	Community-based
Any policy/operational changes at work place, N	32	27
Yes, n (%)	27 (84)	19 (70)
Areas of policy/operational changes, N	27	19
We now actively identify (screen for) malnutrition cases; n (%)	21 (78)	13 (68)
We now diagnose SAM[b] and record in admission and discharge book; n (%)	19 (70)	6 (32)
We now treat SAM cases; n (%)	24 (89)	12 (63)
Non-nutritionists are now able to prepare feeds for children with SAM and do not have to wait for the nutritionist; n (%)	17 (63)	7 (37)
The health facility has provided equipment such as scales, tape measures to enable us to measure children; n (%)	20 (74)	3 (16)

[a] Participant groups at baseline were used for data analysis, and some participants' workplaces may have changed in the follow-up period.

[b] SAM: severe acute malnutrition.

Table 5 presents a summary of operational and policy changes at the participating hospitals and community centers, reported by individual health professionals in Ghana at 12 months. The 2 areas where change was most frequent were in active identification of malnutrition cases at community centers and hospitals and the initiation of treating SAM cases with 2 hospitals in Ghana establishing malnutrition units in the follow-up period.

Changes in Perception

At the 6-month follow-up, participants were asked if there was any change in how they viewed malnutrition and its management as a result of taking the malnutrition eLearning course and if yes, to describe the (most significant) change. Of the 461 participants who responded to the question, 304 said yes and 282 gave a brief description of the change. Table 6 presents a summary of the participants' responses. The most commonly reported change pertained to case-management.

Knowledge Gain, Application, and Confidence in Patient Care Over Time

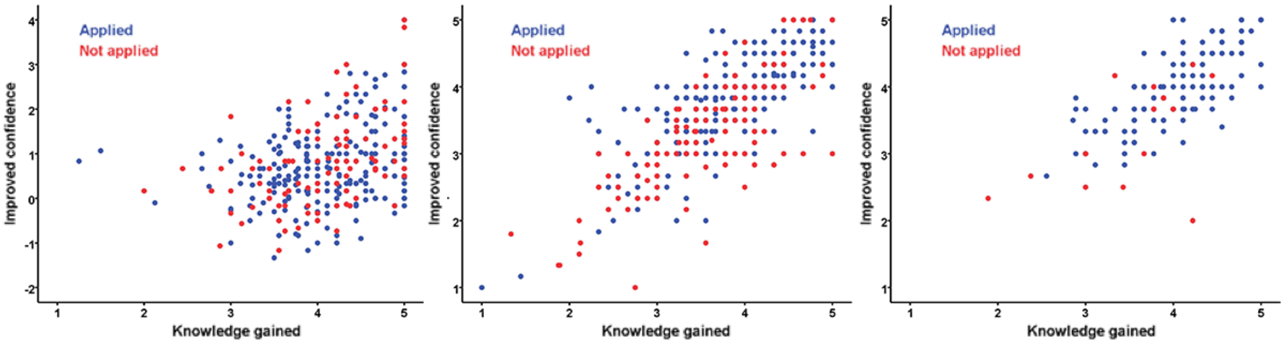
The relation between participants' gain in knowledge, application of knowledge, and confidence in the management of SAM over time was assessed and presented using Spearman rank correlation coefficient (r_s). Figure 4 shows that immediately posttraining, the correlation between the gain in knowledge and confidence was weak, as we had hypothesized. In contrast, a strong correlation emerged at the 6-month follow-up and remained strong at 12 months. The areas where confidence appeared to be most widely strengthened were related to screening children for malnutrition, taking anthropometric measurements, clinical examinations, choosing the correct management option, and managing SAM children using the WHO Ten Steps. These match the areas where the participants applied gained knowledge in patient care (Table 4). Factors that participants reported to have helped improve their knowledge and enable them to apply it in their work were the visual examples in the course, step-by-step descriptions, and practical tasks, all of which made it easier to embrace and follow the concepts of malnutrition and its treatment. Of particular relevance for helping participants gain confidence was the application of the knowledge gained and its use in hands-on experience following completion of the course. The relation between knowledge gain and confidence between 6-month and 12-month follow-up was stronger among the participants who applied knowledge than the ones who did not (applied: N=127, r_s=.74, P<.001 vs not applied: N=18, r_s=.52, P=.03). Important constraining factors for improving confidence were mostly related to the lack of opportunity to apply new knowledge after completion of the course. This appeared especially relevant in institutions where treatment was seen as being mainly the responsibility of doctors.

Table 6. Summary of the reported changes in participants' perception about malnutrition and its management.

Category	Description and example quotes	Total (N=282), n (%)
Perception about malnutrition	Changes in views and perceptions about causes of malnutrition *At first I saw malnutrition as either a curse or it happens to children of parents with low socio-economic background but now I see it in a different light.* [Nutrition trainee, Ghana] *My perception was some child was born with some malnutrition diseases like marasmus. But it has changed because I have learnt in e-learning that is not true.* [Nurse trainee, Ghana] *My perception about SAM has changed greatly and this has increased my knowledge on the causes and prevention of SAM.* [Nurse trainee, Ghana]	55 (19.5)
Assessment of malnutrition	Perception change about physiological and visual characteristics, assessment, and diagnosis of malnutrition *There are visible characteristics that before [we] did not take into account for diagnosing malnutrition.* [Nutrition graduate, Guatemala] *Previously I thought it was only children with severe muscle wasting that were malnourished but I now know that oedematous children are equally malnourished.* [Nurse trainee, Ghana] *It has helped me to differentiate between chronic and acute malnutrition and the signs and symptoms associated with them.* [Principal midwifery officer, Ghana]	51 (18.1)
Management and treatment of SAM	Change in participants' understanding about how and where to treat children with SAM[a] *In relation to indicators and treatment options, that is, if hospital or community.* [Nutrition graduate, El Salvador] *Initially I thought that when stabilize the patient you can just send them home but later I learnt that we have stages. We have the stabilization phase, rehabilitation and recuperation stages, so it changed.* [Health professional, Ghana] *Not all malnourished children should be treated as inpatient. First you need to classify before you start treatment for either in-patient or outpatient.* [Nutrition trainee, Ghana]	127 (45.0)
Professional roles	Perception change about participants' professional role in the management of SAM *Well at first I thought it was only the medical officers who were supposed to treat the medical complications e.g. hypothermia and others but [now] I realize that I could also do it. If not, I could give the instruction for someone to also do it.* [Health professional, Ghana] *First when I go out to work and I see those children, I do not always want to bring them close but with the e-learning I got to know that they also need love and also education.* [Nurse, Ghana]	26 (9.2)
Other	Importance of educating mothers; self-confidence in the management of SAM and eLearning use; views about eLearning *...through the e-learning course I have come to understand that if the mother or caretaker is not well counselled on how to care, prevent and feed the child well, the child's health will not improve.* [Community health worker, Ghana] *E-Learning will help equip nurses and all individuals including myself with the requisite knowledge to detect, manage, educate and treat all SAM cases as I have gained knowledge.* [Nurse trainee, Ghana]	23 (8.2)

[a] SAM: severe acute malnutrition.

Figure 4. The correlation between gain in knowledge and confidence in patient care over time. Improved confidence (y-axis) in plot 1 is calculated by subtracting level of confidence at poststudy from level of confidence at prestudy, and the one in plots 2 and 3 is improved confidence between 2 corresponding time points reported.



1) Between pre- and poststudy N=589, r_s=0.26, P<.001

2) Between post- and 6-month follow-up N=415, r_s=0.76, P<.001

3) Between 6-month and 12-month follow-up N=154, r_s=0.72, P<.001

Discussion

Principal Findings

The goal of the malnutrition eLearning project was to develop an innovative training solution to scale up the training of health professionals in the management of malnutrition. This evaluation study aimed to investigate whether eLearning, designed and developed appropriately, can be an effective means to train health professionals at scale who otherwise may not have the opportunity to receive relevant training. The study has shown that the malnutrition eLearning course is effective in improving the knowledge, understanding, and skills of health professionals in the diagnosis and management of children with SAM, and that the gained knowledge and skills are of practical benefit, enabling health professionals to apply them in their work. Before the training, only 26.3% (94/358) of in-service participants were aware of the WHO management guidelines despite 86.3% (340/394) having regular close involvement in the care of children with SAM. After the training, 85.9% of in-service participants who took the course (128/149) reported applying their new knowledge in their clinical practice, training colleagues who had not participated in the course, and counseling mothers.

The diverse locations and range of health professionals are a unique aspect of our study as most other investigations have been confined to 1 location and a single health profession [57]. For knowledge gained immediately after training and retained knowledge at 6 months, the findings were similar across all contexts. The same findings were observed across different professions. The participants, whether doctors, nutritionists, nurses, or students (preservice), demonstrated that the course enabled them to gain the knowledge, understanding, and skills required for the effective management of SAM. Furthermore, those who had the opportunity to apply their gained knowledge in their daily practice demonstrated higher retained knowledge at 6-month follow-up, corresponding to the theories of learning that knowledge is actively constructed by the learner through mediated or experiential learning [29,51,53]. This appeared to be particularly evident in nurses and midwives who showed the greatest increase between assessment periods, from a relatively low base, which is in accordance with the findings of Murad et al [58].

The training with the malnutrition eLearning course was voluntary and self-directed. This affected the completion of the course, which was set at 3 weeks. At the poststudy, 70.2% (449/640) of the respondents had completed the course, 16.7% (107/640) were in-progress, and 13.1% (84/640) had stopped or not yet started. However, most participants who had partially or fully completed the course actively applied the knowledge and skills they had gained to their clinical practice. They noted opportunities to share their new knowledge with other colleagues who had not participated in the training. Of the participants, 36.8% (146/397) with continued access to the course used it after the training, often together with their colleagues. The trained participants were *empowered* with their gained knowledge and motivated to change their clinical practice applying what they learned. The most cited words in open comments and interviews regarding the most significant benefits they gained from the course were "didn't know but now I know how to...", and "can…" and "am able to...." Positive outcomes encouraged them to seek more opportunities to apply their knowledge with "feeling proud" and "ownership" of the outcomes. Of all the professions, this change was most evident in the nurses and midwives group in Ghana. "Being able to care for SAM children" instead of "having to wait for a nutritionist or a doctor" and "being able to request equipment" with an educated reason were strong motivators for these participants to continue to use their knowledge. Changes initiated by individuals and improved outcomes led to institutional, operational, and policy changes, in contrast to findings from other self-directed learning interventions, which are reported to be effective in the knowledge domain but not in the skills or attitudes domains [58].

Knowledge gained from the course, subsequent application of the knowledge, and experience of positive outcomes changed the perceptions of the participants (66%) about malnutrition and its management, especially in relation to their professional roles in caring for malnourished children. The majority were those who did not have prior training in SAM. As we hypothesized, the relation between gain in knowledge and confidence in patient care was weak immediately after the training but strengthened over time through application of knowledge and experience of positive outcomes, similar to the result observed in an eLearning intervention to improve adherence to guidelines [59]. Learning is not just about increasing one's knowledge, but it is about understanding, seeing things differently, and changing as a person [22-25,60]. The latter aspects of learning are what lead to behavior change, and therefore, they are particularly important to consider when designing and implementing training interventions. The observed changes among the participants suggest that the malnutrition eLearning course and subsequent application of gained knowledge in clinical practice have promoted higher levels of learning, actively advocating the WHO guideline to others, and seeking policy and operational-level changes to improve patient care as well as following the guideline by themselves.

Relevance to Capacity Building

Severe malnutrition in childhood is a complex clinical condition that is challenging to manage but eminently preventable at community level. Under the aegis of WHO and United Nations Children's Fund (UNICEF), the international health community has become highly skilled at effective intervention for prevention and treatment, especially in emergency contexts, but the problem remains a major challenge within the developmental context [1,61]. In large part, this limitation of our ability to apply what is known to be effective on a sustained basis is related to challenges in the education and training of frontline health professionals, especially those in remote, poorly resourced locations with limited support [62]. The WHO Ten Steps have been available for many years, but the treatment of SAM does not feature in the curricula of many medical and nursing schools, with the result that new graduates are ill equipped to care for children with SAM. Added to this is the dearth of tertiary education institutions in low- and middle-income countries that provide training in nutrition. This results in a shortage of trainers and a workforce where only few are competent in the management of severe malnutrition, leading to high mortality among severely ill malnourished children [5]. To address this issue, the international community has developed training materials such as the WHO training course on hospital-based care of SAM [63]. Most, however, are instructor-led, which severely limits the number of participants that can be trained. Available eLearning courses are often knowledge-based and limited to concepts, such as Nutrition in Emergencies by UNICEF [64]. A scalable solution is thus needed that helps the learner gain knowledge and skills and prepares them to improve their practice through self-directed learning.

This study demonstrates that the malnutrition eLearning course offers the opportunity to improve treatment practices of health professionals at scale and to make a significant contribution to building capacity for the care of children with or at risk of severe malnutrition. The internet promotes development through inclusion, efficiency, and innovation and offers great potential to help reduce inequality in training opportunities between developed and developing countries [65]. Where face-to-face training is available, health professionals should take advantage of it. Where face-to-face training is limited, eLearning can provide a useful foundation, freeing-up trainers to spend time on supervisory and support activities.

In this study, we report the process through which participants gained knowledge and their perceptions of how this helped them to change their practice, thereby leading to greater confidence in their abilities to make better-informed judgments and critically reflect upon care options. This learning process may be unremarkable within a well-structured learning environment but is particularly challenging when offered remotely in the context of severely limited resources, where failure in delivering a secure service is not unusual. In a separate paper, we will be reporting that mortality rates among severely malnourished children declined where hospital staff took the malnutrition eLearning course.

Limitations

In this study, we sought to objectively assess the likely benefit of the malnutrition eLearning course with the target groups in locations where the issue of caring for malnourished children is a usual part of the delivery of clinical and health services but the opportunity for training is limited. This in itself is a challenge as the design of a study to achieve these ends is problematic in its execution especially among communities where research is an unusual experience and the imperative of delivering a service is a priority. Therefore, we adopted a pragmatic approach in which participation was voluntary, but the research activities and the execution of those activities were ordered. To obtain some indication of the putative generalizability of the study findings, we included a group of people who chose to take the malnutrition eLearning course for their own reasons online and sought to capture their experience using similar methods.

A weakness of the study is the lower participation rate at 6 months and 12 months. This was largely because in-service professionals who were not on duty or were working in their communities, and the preservice students who were away on placements were unable to take part. Despite the losses, sample sizes were still large enough to show significant differences and the losses were unlikely to create bias. No assessments of remote online users were made at 6 months and 12 months, and they seem to differ from the rest of the subset by having higher baseline scores. We think this might be the result of bias with online users participating in the preassessment because they wanted to test their existing knowledge, rather than advance it.

We chose the Ebel method to assess gain in knowledge. This test is typically used to assess competency, in which case a panel of specifically trained, expert, standard setters is employed. Our team was not equipped to assess competencies, and our study aim was to evaluate whether the malnutrition eLearning course led to quantifiable gains in knowledge and skills. An analysis of the participants' answers to each question, however, may help future setters judge the 3 levels of difficulty utilized in this method. An analysis of answers would also be helpful to identify any aspects of the malnutrition eLearning course that were not well answered so that the questions, or course content, can be improved.

In this study, the malnutrition eLearning course took place without supporting activities. This is unlikely to be the case where the course is being used within a program for scaling up improved management of SAM. For example, in a program for in-service health professionals, one might expect to embed the course within a package of complementary activities such as supervision, job aids, mentoring, audit, and feedback. Such supporting activities would be expected to consolidate learning and result in higher knowledge scores and competency. In this study, assessment of knowledge took place under *exam style* with time limits and no access to manuals, posters, etc, which would be available in ideal workplace situations.

Conclusions

The findings indicate that the self-directed malnutrition eLearning course was well received, and the learning acquired was associated with perceived improvement in practice. The training led to significant changes in knowledge and awareness of best practice in the management of SAM. Importantly, the training encouraged participants to become active change agents, not only changing their own clinical practice by following the WHO guidelines but also promoting changes among their colleagues and caregivers. At some hospitals, the changes made by individuals and associated outcomes led to operational and policy changes including the opening of 2 malnutrition units. Access to IT and internet no longer proved to be a barrier to training.

The approach taken for the capacity development of health professionals in this study can be replicated across different regions and countries; therefore, it should be considered a valuable tool in scaling up the capacity of health professionals in the management of malnourished children. We suggest that the application of good quality eLearning at scale will result in improved knowledge in the treatment and management of SAM, thus providing a scalable mechanism to train high volumes of health professionals and reducing the prevalence and impact of SAM globally.

Acknowledgments

The authors sincerely thank the 12 hospitals, 2 community centers, 9 academic institutions, 2 Ministries of Health, and individual in-service and preservice health professionals who participated in the study. The participating institutions were Maternal and Child Health Hospital, St Michael’s Hospital, Ejura Government Hospital, Konongo Government Hospital, Kumasi South Hospital, Agogo Presby Hospital, St Martin’s Hospital, St Patrick's Hospital, and Mankranso Hospital; Pokukrom and Sabronum Community Health Centres, and the University of Health and Allied Sciences, Kwame Nkrumah University of Science and Technology, Christian Services University College, Kokofu Nurses' Training College, Community Health Nurses Training College, Nursing and Midwifery Training College, and College of Health and Well-being in Ashanti Region, Ghana; and Hospital Juan Pablo II (Guatemala), Hospital Nacional Zacamil (El Salvador), and Hospital Veterinario Universidad de Antioquia (Colombia); Universidad Rafael Landívar and Universidad Evangelica de El Salvador; and the Ministries of Health for Guatemala and El Salvador. The authors greatly value the contributions of those in the United Kingdom, Ghana, and Latin America who enabled and supported the malnutrition eLearning project and its evaluation. The authors would also like to specially thank Sally Ann Hickman for her contributions to this study, assisting the coordination of 3 teams’ activities in the United Kingdom, Ghana, and Latin America, literature search, and the project team workshop.

This report was produced through the support provided by PATH under a grant by the UK Aid from the UK government; however, the views expressed do not necessarily reflect the UK government’s official policies. The funder was not involved in the study design; the collection, analysis, and interpretation of data; the writing of the report; and in the decision to submit the article for publication.

Authors' Contributions

SC and TP conceived and designed the study. RA, MMV, AA, and HMY contributed to the study design. RA and MMV identified the study sites. SC, RA, and MMV supervised and APu, RA, MMV, NELA, SKB, CIVM, CEPS, and APe carried out the data collection. SC, HMY, APe, MMV, NELA, and SKB performed data cleaning. SC, HMY, and APu analyzed and interpreted the data, with HMY carrying out statistical analysis and SC and APu qualitative data analysis. SC led the preparation of the manuscript. HMY and TP contributed to the manuscript preparation, and RA, MMV, APu, and APe reviewed the script. AAJ and AA critically reviewed and revised the manuscript. All authors participated in the final approval of the manuscript. All authors had full access to all of the data (including statistical reports and tables) in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis.

Conflicts of Interest

AA, RA, SC, and AAJ were part of the malnutrition eLearning development team.

Multimedia Appendix 1

Malnutrition eLearning-background and goal.

[PDF File (Adobe PDF File), 9MB - jmir_v20i10e10396_app1.pdf]

Multimedia Appendix 2

Results and findings from a field test in Uganda.

[PNG File, 819KB - jmir_v20i10e10396_app2.png]

Multimedia Appendix 3

Reported changes made in clinical practice by individuals at 12-month follow-up.

[PDF File (Adobe PDF File), 59KB - jmir_v20i10e10396_app3.pdf]

References

1. Black RE, Victora CG, Walker SP, Bhutta ZA, Christian P, de Onis M, et al. Maternal and child undernutrition and overweight in low-income and middle-income countries. The Lancet 2013 Aug;382(9890):427-451. [doi: 10.1016/S0140-6736(13)60937-X] [Medline: 23746772]
2. Bhutta ZA, Das JK, Rizvi A, Gaffey MF, Walker N, Horton S, Lancet Nutrition Interventions Review Group, the Maternal and Child Nutrition Study Group. Evidence-based interventions for improvement of maternal and child nutrition: what can be done and at what cost? Lancet 2013 Aug 03;382(9890):452-477. [doi: 10.1016/S0140-6736(13)60996-4] [Medline: 23746776]
3. Jackson AA, Ashworth A, Mokhtar N, Uauy R. Severe malnutrition: building on the past for a brighter future. Food Nutr Bull 2014 Jun;35(2 Suppl):S3-S9. [doi: 10.1177/15648265140352S101] [Medline: 25069287]
4. Gillespie S, Haddad L, Mannar V, Menon P, Nisbett N, Maternal and Child Nutrition Study Group. The politics of reducing malnutrition: building commitment and accelerating progress. Lancet 2013 Aug 10;382(9891):552-569. [doi: 10.1016/S0140-6736(13)60842-9] [Medline: 23746781]
5. Shrimpton R, Hughes R, Recine E, Mason JB, Sanders D, Marks GC, et al. Nutrition capacity development: a practice framework. Public Health Nutr 2014 Mar;17(3):682-688. [doi: 10.1017/S1368980013001213] [Medline: 23651529]
6. Schofield C, Ashworth A, Annan R, Jackson AA. Malnutrition treatment to become a core competency. Arch Dis Child 2012 May;97(5):468-469. [doi: 10.1136/adc.2010.209015] [Medline: 21427122]
7. Hossain M, Chisti MJ, Hossain MI, Mahfuz M, Islam MM, Ahmed T. Efficacy of World Health Organization guideline in facility-based reduction of mortality in severely malnourished children from low and middle income countries: a systematic review and meta-analysis. J Paediatr Child Health 2017 May;53(5):474-479. [doi: 10.1111/jpc.13443] [Medline: 28052519]
8. Karaolis N, Jackson D, Ashworth A, Sanders D, Sogaula N, McCoy D, et al. WHO guidelines for severe malnutrition: are they feasible in rural African hospitals? Arch Dis Child 2007 Mar;92(3):198-204 [FREE Full text] [doi: 10.1136/adc.2005.087346] [Medline: 16670119]
9. Landi M, Swakin E, Minijihau C, Welch H, Tefuarani N, Duke T. Severe malnutrition in children in Papua New Guinea: effect of a multi-faceted intervention to improve quality of care and nutritional outcomes. Paediatr Int Child Health 2017 Feb;37(1):21-28. [doi: 10.1080/20469047.2015.1106079] [Medline: 26750043]
10. Puoane T, Cuming K, Sanders D, Ashworth A. Why do some hospitals achieve better care of severely malnourished children than others? Five-year follow-up of rural hospitals in Eastern Cape, South Africa. Health Policy Plan 2008 Nov;23(6):428-437. [doi: 10.1093/heapol/czn036] [Medline: 18796499]
11. Jackson AA, Ashworth A, Khanum S. Improving child survival: Malnutrition Task Force and the paediatrician's responsibility. Arch Dis Child 2006 Aug;91(8):706-710 [FREE Full text] [doi: 10.1136/adc.2006.095596] [Medline: 16861489]
12. Mucha N, Tharaney M. Ministry of Health and Family Welfare.: Bread for the World Institute/Helen Keller International; 2013. Strengthening human capacity to scale up nutrition URL: http://www.a2hp.mohfw.gov.in/ResourceDocuments/Strengthening%20Human%20Capacity%20to%20scape%20up%20Nutrition.pdf [WebCite Cache ID 72gAk0TrL]
13. Geissler C. Capacity building in public health nutrition. Proc Nutr Soc 2015 Nov;74(4):430-436. [doi: 10.1017/S0029665114001736] [Medline: 25604975]
14. Fanzo JC, Graziose MM, Kraemer K, Gillespie S, Johnston JL, de Pee S, et al. Educating and training a workforce for nutrition in a post-2015 world. Adv Nutr 2015 Nov;6(6):639-647 [FREE Full text] [doi: 10.3945/an.115.010041] [Medline: 26567189]
15. Al-Shorbaji N, Atun R, Car J, Majeed A, Wheeler E. eLearning for undergraduate health professional education: a systematic review informing a radical transformation of health workforce development. Geneva: World Health Organization; 2015.
16. Sinclair PM, Kable A, Levett-Jones T, Booth D. The effectiveness of Internet-based e-learning on clinician behaviour and patient outcomes: a systematic review. Int J Nurs Stud 2016 May;57:70-81. [doi: 10.1016/j.ijnurstu.2016.01.011] [Medline: 27045566]
17. Voutilainen A, Saaranen T, Sormunen M. Conventional vs. e-learning in nursing education: a systematic review and meta-analysis. Nurse Educ Today 2017 Mar;50:97-103. [doi: 10.1016/j.nedt.2016.12.020] [Medline: 28038371]
18. Sissine M, Segan R, Taylor M, Jefferson B, Borrelli A, Koehler M, et al. Cost comparison model: blended eLearning versus traditional training of community health workers. Online J Public Health Inform 2014;6(3):e196 [FREE Full text] [doi: 10.5210/ojphi.v6i3.5533] [Medline: 25598868]
19. Piaget J. Piaget's theory. In: Mussen PH, editor. Carmichael's Manual of child psychology, volume 1. 3rd ed. New York: John Wiley & Sons Inc; 1970.

20. Duffy TM, Jonassen DH, editors. Constructivism and the technology of instruction: a conversation. New Jersey: Routledge; 1992.
21. Duffy TM, Cunningham DJ. Constructivism: implications for the design and delivery of instruction. In: Janassen DH, editor. Handbook of research for educational communications technology. New York: Simon and Schuster; 1996:170-198.
22. Marton F, Dall'Alba G, Beatty E. Conceptions of learning. IJER 1993;19:277-300.
23. Marton F, Booth S. Learning and awareness. New Jersey: Lawrence Erlbaum Associates; 1997.
24. Prosser M, Trigwell K. Understanding learning and teaching. Buckingham [England]: Society for Research into Higher Education & Open University Press; 1999.
25. Ramsden P. Learning to teach in higher education. 2nd edition. London: Routledge; 2003.
26. Marton F, Saljo R. Approaches to learning. In: Marton F, Hounsell D, Entwistle N, editors. The experience of learning: implications for teaching and studying in higher education. 3rd (internet) edition. Edinburgh: University of Edinburgh, Centre for Teaching, Learning and Assessment; 2005:39-58.
27. Marton F, Ramsden P. What does it take to improve learning? In: Ramsden P, editor. Improving learning: new perspectives. London: Kogan Page; 1988.
28. Marouchou DV. Can students' concept of learning influence their learning outcomes? High Learn Res Commun 2012 Jun 01;2(2):18-33. [doi: 10.18870/hlrc.v2i2.23]
29. Laurillard D. Rethinking university teaching: a conversational framework for the effective use of learning technologies. 2nd ed. London: Routledge/Falmer; 2002.
30. Wood DF. Problem based learning. Br Med J 2008 May 03;336(7651):971 [FREE Full text] [doi: 10.1136/bmj.39546.716053.80] [Medline: 18456594]
31. Anderson JR, Reder LM, Simon HA. Situated learning and education. ER 1996 May;25(4):5-11. [doi: 10.3102/0013189X025004005]
32. Choi S, Webb A, Heyworth J, Golestani F, Slaght S. Virtual patients: year 1. MEDEV Newsletter 2010;22(1):19-20 ISSN 1479-523X (online).
33. Ellaway RH, Poulton T, Smothers V, Greene P. Virtual patients come of age. Med Teach 2009 Sep 09;31(8):683-684. [doi: 10.1080/01421590903124765]
34. Cook DA, Triola MM. Virtual patients: a critical literature review and proposed next steps. Med Educ 2009 Apr;43(4):303-311. [doi: 10.1111/j.1365-2923.2008.03286.x] [Medline: 19335571]
35. Cook DA, Erwin PJ, Triola MM. Computerized virtual patients in health professions education: a systematic review and meta-analysis. Acad Med 2010 Oct;85(10):1589-1602. [doi: 10.1097/ACM.0b013e3181edfe13] [Medline: 20703150]
36. Saleh N. The value of virtual patients in medical education. Ann Behav Sci Med Educ 2015 Oct 16;16(2):29-31. [doi: 10.1007/BF03355129]
37. Consorti F, Mancuso R, Nocioni M, Piccolo A. Efficacy of virtual patients in medical education: a meta-analysis of randomized studies. Comput Educ 2012 Nov;59(3):1001-1008. [doi: 10.1016/j.compedu.2012.04.017]
38. Fontaine G, Cossette S, Maheu-Cadotte M, Mailhot T, Deschênes MF, Mathieu-Dupuis G. Effectiveness of adaptive e-learning environments on knowledge, competence, and behavior in health professionals and students: protocol for a systematic review and meta-analysis. JMIR Res Protoc 2017 Jul 05;6(7):e128 [FREE Full text] [doi: 10.2196/resprot.8085] [Medline: 28679491]
39. Yang TC, Hwang GJ, Yang SJ. Development of an adaptive learning system with multiple perspectives based on students' learning styles and cognitive styles. JETS 2013;16(4):185-200 ISSN 1176-3647 e-ISSN 1176-3647.
40. Pickup T, Annan R, Choi S. Malnutrition eLearning course. African Nutrition Matters 2014 Jun;2(2):13-14 [FREE Full text]
41. Kirkpatrick Partners. The Kirkpatrick Model URL: https://www.kirkpatrickpartners.com/Our-Philosophy/The-Kirkpatrick-Model [WebCite Cache ID 6xmw0uZKX]
42. UNICEF, World Health Organization, The World Bank Group. World Health Organization. 2017. Joint child malnutrition estimates-levels and trends in child malnutrition (2017 edition) URL: http://www.who.int/nutgrowthdb/jme_brochoure2017.pdf?ua=1 [WebCite Cache ID 6xosCSVs2]
43. Choi S, Annan R, Jackson AA. eLearning as a solution to train doctors and nurses to manage malnutrition in children. In: Proceedings of CAPGAN Congress. GAPGAN. 2011. Presented at: CAPGAN Congress; 2011 July 21-23; London: UK.
44. Ashworth A, Khanum S, Jackson AA, Schofield C. Guidelines for the inpatient treatment of severely malnourished children. Geneva: World Health Organization; 2003.
45. World Health Organization. Pocket book of hospital care for children: guidelines for the management of common childhood illnesses. 2nd ed. Geneva: World Health Organization; 2013.
46. World Health Organization. Guideline: updates on the management of severe acute malnutrition in infants and children. Geneva: World Health Organization; 2013.
47. University of Southampton. som.soton. 2011. Nutrition eLearning portal welcome page URL: http://www.som.soton.ac.uk/learn/test/nutrition/ [accessed 2018-09-24] [WebCite Cache ID 6tIq15bK3]
48. University of Southampton. som.soton. 2011. We eLearn Do You? (WeNDY) URL: http://www.som.soton.ac.uk/learn/elearning/events/ [accessed 2018-09-24] [WebCite Cache ID 72Kw4O8Rd]

49. Ke S, Choi S, Ault E. Technical solutions for competency-based eLearning for public health nutrition. In: Herrington J, Montgomerie C, editors. Proceedings of ED-MEDIA 2010-World Conference on Educational Multimedia, Hypermedia & Telecommunications. Toronto: Association for the Advancement of Computing in Education; 2010:3886-3895.
50. Hammerton M, Choi S, Ke G, Livy A, Duboc J. Improving efficiency of bespoke eLearning resource development with reusable ActionScript 3 classes. In: Herrington J, Montgomerie C, editors. Proceedings of ED-MEDIA 2010-World Conference on Educational Multimedia, Hypermedia & Telecommunications. Toronto: Association for the Advancement of Computing in Education; 2010:3040-3045.
51. Kolb DA. Experiential learning: experience as the source for learning and development. 2nd ed. New Jersey: Person Education Inc; 2015.
52. Laurillard D, Stratfold M, Luckin R, Plowman L, Taylor J. Affordances for learning in a non-linear narrative medium. JIME 2000(2):p.Art.2. [doi: 10.5334/2000-2]
53. Mayes JT, Fowler CJ. Learning technology and usability: a framework for understanding courseware. Interacting with Computers 1999 May;11(5):485-497. [doi: 10.1016/S0953-5438(98)00065-4]
54. Ebel R. Essential of Educational Measurement. Englewood Cliffs, NJ: Prentice-Hall; 1972.
55. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
56. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960 Apr;20(1):37-46. [doi: 10.1177/001316446002000104]
57. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Internet-based learning in the health professions: a meta-analysis. J Am Med Assoc 2008 Sep;300(10):1181-1196. [doi: 10.1001/jama.300.10.1181] [Medline: 18780847]
58. Murad MH, Coto-Yglesias F, Varkey P, Prokop LJ, Murad AL. The effectiveness of self-directed learning in health professions education: a systematic review. Med Educ 2010 Nov;44(11):1057-1068. [doi: 10.1111/j.1365-2923.2010.03750.x] [Medline: 20946476]
59. de Beurs DP, de Groot MH, de Keijser J, Mokkenstorm J, van Duijn E, de Winter RF, et al. The effect of an e-learning supported Train-the-Trainer programme on implementation of suicide guidelines in mental health care. J Affect Disord 2015 Apr;175:446-453 [FREE Full text] [doi: 10.1016/j.jad.2015.01.046] [Medline: 25679199]
60. Saljo R. Learning about learning. High Educ 1979 Jul;8(4):443-451. [doi: 10.1007/BF01680533]
61. WHO, UNICEF, WFP, UN System Standing Committee on Nutrition. Community-based management of severe acute malnutrition. Geneva: WHO; 2007.
62. Sodjinou R, Bosu WK, Fanou N, Déart L, Kupka R, Tchibindat F, et al. Nutrition training in medical and other health professional schools in West Africa: the need to improve current approaches and enhance training effectiveness. Global Health Action 2014 Jul 30;7(1):24827. [doi: 10.3402/gha.v7.24827]
63. World Health Organization. Geneva: World Health Organization; 2009. Training course on the management of severe malnutrition URL: http://www.who.int/nutrition/publications/severemalnutrition/training_inpatient_MSM/en/ [accessed 2018-09-24] [WebCite Cache ID 72L2jjrlz]
64. UNICEF. 2010. Nutrition in emergencies URL: http://www2.unicef.org/nutrition/training/ [WebCite Cache ID 72L37Jerc]
65. World Bank Group. World development report 2016: digital dividends. Washington DC: International Bank for Reconstruction and Development/The World Bank; 2016.

Abbreviations

IT: information technology
SAM: severe acute malnutrition
UNICEF: United Nations Children's Fund
WHO: World Health Organization

Edited by G Eysenbach; submitted 13.03.18; peer-reviewed by S Edelbring, G Fontaine; comments to author 19.04.18; revised version received 11.09.18; accepted 14.09.18; published 03.10.18.

Please cite as:
Choi S, Yuen HM, Annan R, Pickup T, Pulman A, Monroy-Valle M, Aduku NEL, Kyei-Boateng S, Velásquez Monzón CI, Portillo Sermeño CE, Penn A, Ashworth A, Jackson AA
Effectiveness of the Malnutrition eLearning Course for Global Capacity Building in the Management of Malnutrition: Cross-Country Interrupted Time-Series Study
J Med Internet Res 2018;20(10):e10396
URL: http://www.jmir.org/2018/10/e10396/
doi:10.2196/10396
PMID:

Original Paper

A Mobile Game for Patients With Breast Cancer for Chemotherapy Self-Management and Quality-of-Life Improvement: Randomized Controlled Trial

Hee Jun Kim[1], MD, PhD; Sun Mi Kim[2], MD, PhD; Heechul Shin[3], MD, PhD; Joung-Soon Jang[1], MD, PhD; Young In Kim[2], MD; Doug Hyun Han[2], MD, PhD

[1]Department of Internal Medicine, College of Medicine, Chung-Ang University, Seoul, Republic Of Korea
[2]Department of Psychiatry, College of Medicine, Chung-Ang University, Seoul, Republic Of Korea
[3]Department of Surgery, College of Medicine, Chung-Ang University, Seoul, Republic Of Korea

Corresponding Author:
Doug Hyun Han, MD, PhD
Department of Psychiatry
College of Medicine
Chung-Ang University
224-1 Heukseok-dong
Dongjak-gu
Seoul,
Republic Of Korea
Phone: 82 2 6299 3132
Fax: 82 2 6299 2064
Email: hduk70@gmail.com

Abstract

Background: The application of game-based learning in clinical practice has shown potential advantages in previous studies. However, there have been little efforts to use smartphone-based mobile games in the management of adult patients with cancer.

Objective: The objective of our study was to evaluate if patient education using a mobile game may increase drug compliance, decrease physical side effects of chemotherapy, and improve psychological status in breast cancer patients.

Methods: A total of 76 patients with metastatic breast cancer who were planned to receive cytotoxic chemotherapy were enrolled in this trial. Study participants were randomly assigned to a mobile game play group (game group, n=36) or a conventional education group (control group, n=40) in a ratio of 1:1. The patients were unblinded and followed prospectively for 3 weeks. Outcome measures included time spent for education, compliance to medication, physical side effects, and psychological side effects including quality of life (QoL).

Results: Overall, 72 out of 76 patients completed the study after 3 weeks (95%). The subjects in the game group showed high levels of satisfaction with the app. The time spent playing the mobile game in the game group was longer than that spent for self-education in the control group (mean 22.2, SD 6.1 vs mean 5.5, SD 4.0 minutes a day; *P*<.001). The mobile game group showed better drug adherence (Korean version of the Medication Adherence Rating Scale; mean 7.6, SD 0.7 vs mean 6.5, SD 0.5; *P*<.001). The use of the mobile game was associated with lower rates of chemotherapy-related side effects, such as nausea, fatigue, numbness of hand or foot, and hair loss, than the control group. The game group exhibited better QoL during chemotherapy (mean 74.9, SD 3.5 vs mean 72.2, SD 5.3; *P*=.01). However, there were no significant differences in terms of depression and anxiety scales.

Conclusions: This study suggests the feasibility and potentiality of the use of smartphone mobile games for patients with breast cancer receiving chemotherapy. Education using a mobile game led to better patient education, improved drug compliance, decreased side effects, and better QoL compared with conventional education. Mobile games can be used as easy, fun, and effective measures for patient education and have the potential to improve treatment outcomes.

Trial Registration: ClinicalTrials.gov NCT03205969; http://clinicaltrials.gov/ct2/show/NCT03205969 (Archived by WebCite at http://www.webcitation.org/71jfSBOq9).

(J Med Internet Res 2018;20(10):e273) doi:10.2196/jmir.9559

KEYWORDS

mobile phone; breast cancer; chemotherapy; side effects; quality of life

Introduction

Playing console and Web-based games is a popular free-time activity among children, adolescents, and adults [1]. Recent studies with health-related internet games have shown positive effects, such as improving coping strategies for health problems, strengthening treatment compliance, and increasing motivation to overcome the difficult times during illness [2]. Potential benefits of using computer-based patient education programs include increased patient knowledge in perceived information competence [3]. Game-based learning may be enjoyable, interesting, and immersive and therefore, more effective than classical learning [4]. Patient education has shown to be effective in the management of patients with teenage cancer, diabetes, and asthma [2]. However, there have been limited studies conducted in adults.

Chemotherapy is the main treatment for breast cancer, which has proven to reduce the rate of recurrence and mortality in breast cancer patients [5]. Chemotherapy is accompanied by significant side effects. Consequently, many cancer survivors experience physical and psychological symptoms that hamper their quality of life (QoL) and disrupt their daily living activities, family relationships, and work schedules [6]. Diarrhea, nausea, vomiting, hair loss, and mucositis are among the most common side effects [7]. These side effects may cause poor drug compliance, prohibiting successful anticancer treatment. Poor education is one of the main determinants of poor adherence to chemotherapy [8]. Therefore, proper education and sufficient communication are important to increase adherence, which may eventually contribute to improved clinical outcomes [9,10].

An easily accessible, immersive, and interactive Web-based game (ILOVEBREAST) was developed for adult patients with metastatic breast cancer. Although previous efforts using games were mostly focused on children, adolescents, or young adults, ILOVEBREAST is specifically designed for adult users. This study was a proof-of-concept randomized controlled trial aimed at evaluating the benefits of smartphone-based mobile game use in breast cancer patients receiving cytotoxic chemotherapy. We hypothesized that mobile gaming would lead to increased drug compliance, decreased physical side effects of chemotherapy, and improved psychological status among patients.

Methods

Patients

Patients with pathologically proven, clinical stage IV breast cancer were enrolled in this study at Chung-Ang University Hospital, Korea, from September 2013 to September 2014. Patients who had metastatic breast cancer and agreed to participate in an education-controlled trial of mobile game management were screened with the research and the patient edition version of Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision Axis I Disorders [11]. The inclusion criteria were as follows: females diagnosed with metastatic breast cancer, aged 18-65 years, use of at least third-line palliative chemotherapy treatment comprising taxanes, anthracyclines, capecitabine, and platinum compounds, and ability to use a smartphone for the mobile game. The exclusion criteria were as follows: current or history of uncontrolled medical diseases except for breast cancer, psychiatric diseases, including major depressive disorder and anxiety disorders, and history of substance abuse, including alcohol, nicotine, and drugs. Major depressive disorder and anxiety disorder were diagnosed based on the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision, the authoritative guide to the diagnosis of mental disorders published by American Psychiatric Association [12].

The Chung-Ang University Hospital Institutional Review Board approved the research protocol for this study (Number C20141447). Informed consent was obtained from all patients during hospitalization for chemotherapy after explaining the design, protocol, and consequences of the study (NCT03205969).

Description of the Mobile Game

A mobile game, ILOVEBREAST (CLGAMES, Seoul, Korea), was developed with an intention to improve self-management and to reduce the side effects of chemotherapy drugs. The ILOVEBREAST program used in this study was a 3-week program using typical multiplayer, social network, and platform-based features. The game's key pedagogical features were as follows: education for preventing side effects of anticancer drugs, support for the prevention of side effects of anticancer drugs including numbness, hair loss, and loss of appetite, encouragement of mood and activity, including exercise, pet walking, cooking, and social game playing, which may facilitate participation in such activities in real life, and self-assessment using a personal avatar, as seen in Figure 1.

At the start of game play, an avatar is generated based on the patient's medical information including the status of blood components, general medical condition, and chemotherapeutic drugs. Avatars are to visit their home, pharmacy, hospital, and gymnasium; make purchases from shops; and operate a farm to harvest ingredients for food. They can receive prescribed medications, cook food appropriate for their health, and exercise. Depending on the patient's medication dosage, avatars can pursue a quest to minimize the side effects. The quest consists of taking the medication at the right time, cooking a meal for oneself, exercising, going outside for a walk, and chatting with a friend. An alarm alerts the avatar to take medications timely. Each time the avatar accomplishes a quest, "heart coins" are rewarded. The greater the number of the coins the patient receives, the greater the improvement in the avatar's health status. The avatar can then use these coins to purchase items such as hats, gloves, and food ingredients. Patients are asked to visit their doctor weekly to assess their health status.

Figure 1. Representative screenshots of the ILOVEBREAST game. Source and copyright: Industry Academic Cooperation Foundation, Chung-Ang University Hospital.

Study Procedure

This study was a 3-week prospective trial. All study participants received a combination of 4 chemotherapy drugs (taxanes, anthracyclines, capecitabine, and cisplatin). Patients were randomly assigned to education using mobile game play (game group) or conventional education (control group) utilizing an interactive Web randomization system.

For patients in the game group, the study mobile game (ILOVEBREAST) was installed on the participants' smartphones. They were recommended to play the game for >30 minutes a day, 3 times per week. They were interviewed every week via cell phone until the end of the study. The patients in the game group were checked with the record of access to the ILOVEBREAST game. The patients in the control group received routine care. A brochure elaborating the coping strategies for the side effects of chemotherapy drugs was provided to the control group only. The 26-page education material consisted of 2 parts. Part 1 provided the overall guidelines relating to life patterns, definition and purpose of chemotherapy, and broad side effects of chemotherapy. Part 2 contained individualized educational material comprising each patient's purpose for chemotherapy, the name of the anticancer drugs, chemotherapy schedule, individualized management of side effects and symptoms, and guidance for daily life and mental attitude. The subjects in the control group were recommended to read the material for >30 minutes a day, 3 times per week. The patients in the game group were requested to rate the level of satisfaction using the following 8 questions:

1. What percentage of game content do you use while playing the game?
2. Is the game difficult to play?
3. Is the game fun?
4. Is the game helpful for taking your medication?
5. Does the game provide you with information about breast cancer and treatment?
6. Does the game decrease your unease with chemotherapy?
7. Do you plan to play ILOVEBREAST during your next chemotherapy session?
8. Would you recommend ILOVEBREAST to other patients with breast cancer?

The questions were assessed using a self-reported scale of 10 levels with 10 indicating “very bad” to 100 indicating “very good.”

Trial outcomes were measured in the following 4 domains: time spent for education, compliance to medication, physical side effects, and psychological side effects including QoL. Education time was measured as either the time spent for game playing or self-education using the brochure with preventive measures. The manufacturer (CLGAMES, Seoul, Korea) provided the time the users spent playing the mobile game. Medication compliance was assessed with the Korean version of the Medication Adherence Rating Scale (K-MARS), which has a Cronbach reliability alpha of .71 [13].

Physical and psychological side effects were assessed at baseline and the end of the 3-week follow-up period using questionnaires [7]. The questionnaires for the presence and severity of side effects (ie, nausea, fatigue, decreased appetite, numbness on hand or foot, stomatitis, diarrhea or constipation, hair loss, and skin rash) were measured using a 5-point Likert scale. Questionnaires for psychological assessment included the Beck

Depression Inventory (BDI) with a Cronbach reliability alpha of .83 [14], the Spielberger State-Trait Anxiety Scale with alpha of .84 [15], and the World Health Organization Quality of Life-BREF Scale with alpha of .87 [16]. BDI is one of the most widely used self-reported scales to assess the severity of depressive mood and consists of 21 multiple-choice [14] questions. The Spielberger State-Trait Anxiety Scale has 20 items each to measure trait anxiety and state anxiety [17]. State anxiety means a temporary response to perceived threats, whereas trait anxiety refers to a consistent personality trait to experience anxiety. For this study, only state anxiety was measured because we intended to measure short-term response related to chemotherapy [15]. The World Health Organization Quality of Life-BREF Scale is an abbreviated generic Quality of Life Scale developed through the World Health Organization and consists of 26 items in the following 4 domains: physical health, psychosocial health, social relationships, and environment [16].

Statistics

Because this study was a proof-of-concept trial, the sample size calculation was done on a practical basis. In the center where this study was undertaken, an average of 200 patients receives cytotoxic chemotherapy in a month. We assumed approximately 30 patients a month would meet the inclusion and exclusion criteria of this study and planned to recruit 90 patients for 3 months. Patient enrollment was slower than expected, and we decided to stop enrollment 1 year after the initiation of the study.

Continuous variables were presented as mean (SD), and categorical variables were presented as counts and percentages. Continuous variables were compared with independent *t* tests or Mann-Whitney U-tests as appropriate. The Chi-square test or Fisher's exact test was used for dichotomous variables. For all statistical analyses, the significance level alpha was set at.05, and all analyses were performed using SPSS 18.0 (Chicago, IL, USA).

Results

Demographic Characteristics

A flow diagram of the study is shown in Figure 2. A total of 83 patients with metastatic breast cancer agreed to participate in the study. Among them, 2 were excluded because of severe depressive and anxiety symptoms, 4 for having difficulties in using the mobile game, and 1 for withdrawal of consent for an unspecified reason. Among the 76 female patients who were finally enrolled, 36 and 40 were randomly assigned to the mobile game and control groups, respectively. After 3 weeks of the study duration, 34 and 38 patients in each group completed the study (94.7%).

The baseline characteristics of the study subjects are summarized in Table 1. The mean age was 50.9 (SD 7.0) years, and all the patients were female. All the participants had an Eastern Cooperative Oncology Group performance status between 0 and 2. There were no significant differences in demographics status and socioeconomic status, pathologic classification, and hormonal status between the 2 groups.

Overall Satisfaction

The patients in the game group were requested to assess their level of satisfaction with ILOVEBREAST, as seen in Figure 3. Patients in the game group played 41% of the game contents (quests, level ups, and rewards), and more than half responded that it was difficult to use (20/36, 56%). However, 67% (24/36) and 61% (22/36) of the patients responded that the game was fun and helpful in taking their medications, respectively. Approximately ¾ of patients (27/36) responded positively in terms of acquisition of information (74.4%), usefulness in overcoming chemotherapy side effects (73.9%), and willingness to play again (72.2%). Furthermore, approximately 81% reported they would recommend the game to other patients with breast cancer.

Figure 2. CONSORT study flow diagram.

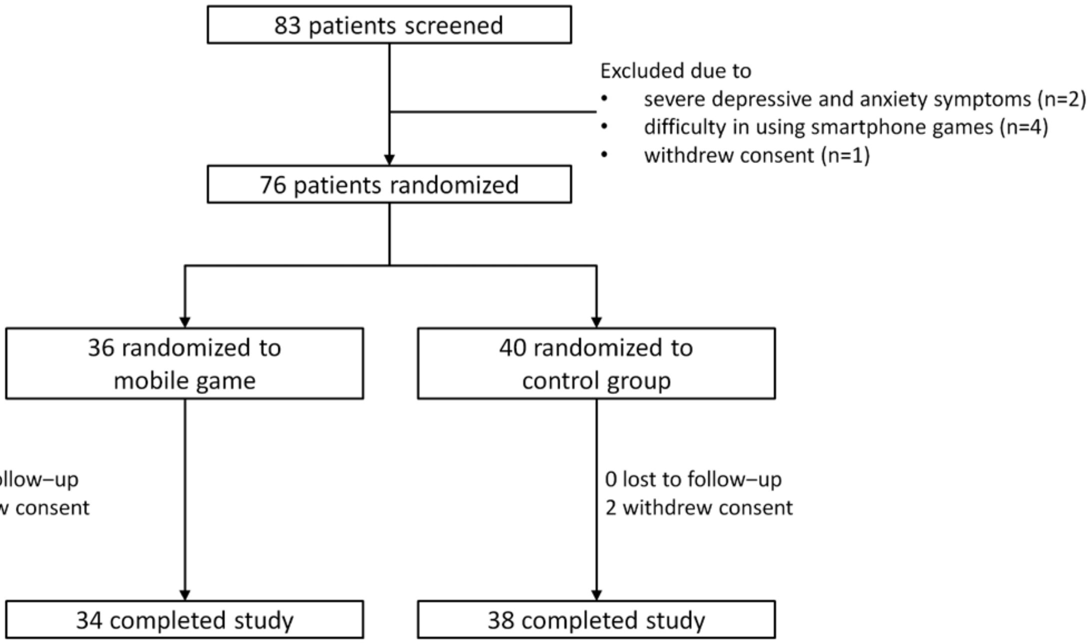


Table 1. Demographic data.

Characteristics	Game group (n=36)	Control group (n=40)	*P* value
Age (years), median	49.8	52.1	.24
Years of education, mean (SD)	13.5 (2.0)	13.2 (1.9)	.51
Economic status[a], n (%)			.94
Highest tertile	2 (6)	2 (5)	
Middle tertile	32 (89)	35 (88)	
Lowest tertile	2 (6)	3 (8)	
Current smoker, n (%)	5 (14)	4 (10)	.43
Social drinking, n (%)	17 (47)	19 (48)	.49
History of mobile gaming, n (%)	20 (56)	22 (55.0)	.50
Performance status, n (%)			.94
ECOG[b] 0-1	35 (97)	39 (98)	
ECOG 2	1 (3)	1 (2.5)	
ECOG 3-4	0 (0)	0 (0.0)	
Pathologic characteristics, n (%)			.99
Invasive ductal carcinoma	35 (97)	38 (95)	
Others	1 (3)	2 (5)	
Stage, n (%)			.79
Initial stage IV	8 (22)	10 (25)	
Relapsed stage IV	26 (72)	30 (75)	
Hormone receptor status, n (%)			.77
Positive	25 (69)	29 (73)	
Negative	11 (31)	11 (28)	
Human epidermal growth factor receptor 2, n (%)			.07
Positive	13 (36)	7 (18)	
Negative	23 (64)	33 (83)	
Triple negative phenotype, n (%)			.84
Nontriple negative	30 (83)	34 (85)	
Triple negative	6 (17)	6 (15)	

[a] Economic status was classified into the following tertiles: highest, > US $100,000; middle, $30000-$100,000; and lowest, <$30000.

[b] ECOG: Eastern Cooperative Oncology Group.

Study Endpoints

The time spent on game playing in the mobile game group was higher than that spent for self-education in the control group (22.2, SD 6.1 vs 5.5, SD 4.0 minutes; *P*<.001; Table 2). The game group also showed improved compliance to medications compared with the control group (K-MARS score, 7.6, SD 0.7 vs 6.5, SD 0.5; *P*<.001). The patients in the study group reported lower rates of physically adverse events, such as nausea (*P*=.02), fatigue (*P*=.02), and numbness in the hand or foot (*P*=.02). Clinically significant adverse events, defined by grade ≥3 of Common Terminology Criteria for Adverse Events 3.0, including nausea (*P*=.02), fatigue (*P*=.002), and hair loss (*P*=.01) was shown to be lower in the game group.

After the study period of 3 weeks, the game group showed a higher QoL than in the control group, as seen in Table 3 and Figure 4. The decrease in the QoL score was also significantly lower in the game group than in the control group (*P*=.01). Regarding the subitems of QoL, the use of the mobile game was associated with lower decreases in physical health and environment but a higher decrease in psychological health than the usual care. There were no significant differences in the BDI score or state anxiety scale between the 2 groups.

Figure 3. Levels of satisfaction in ILOVEBREAST.

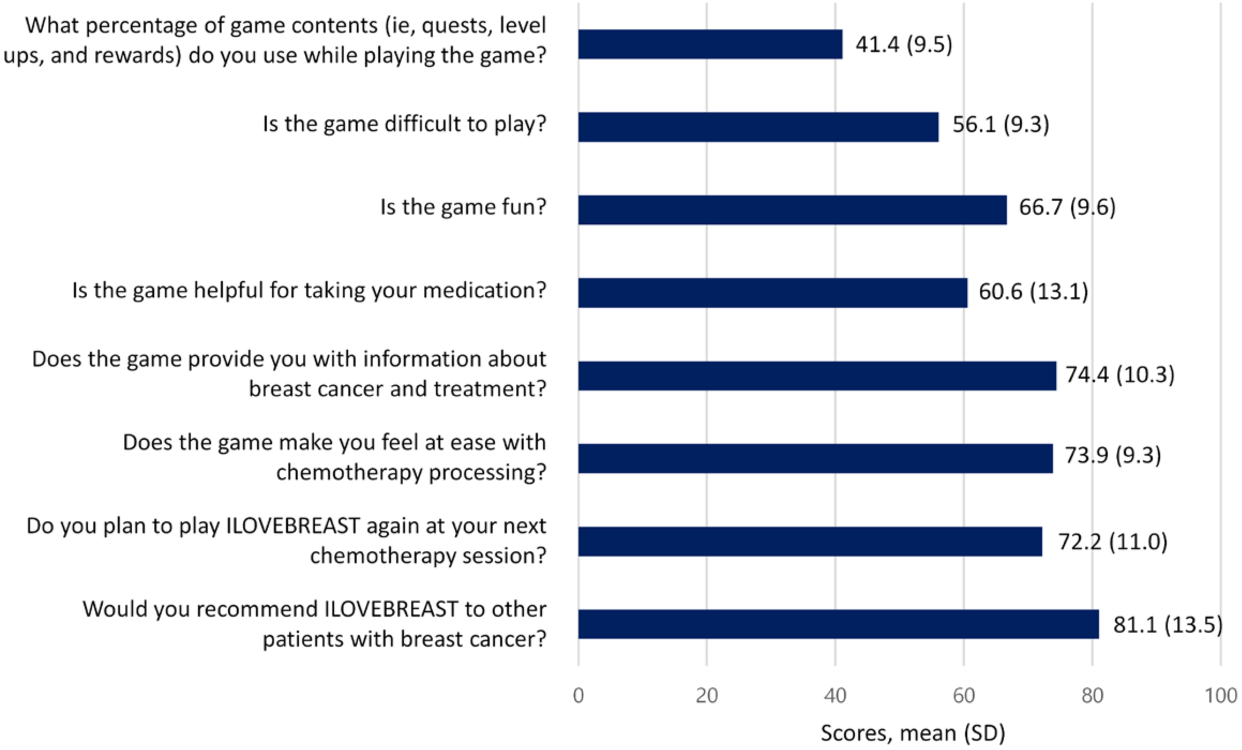


Table 2. Comparison of physically adverse events between study groups.

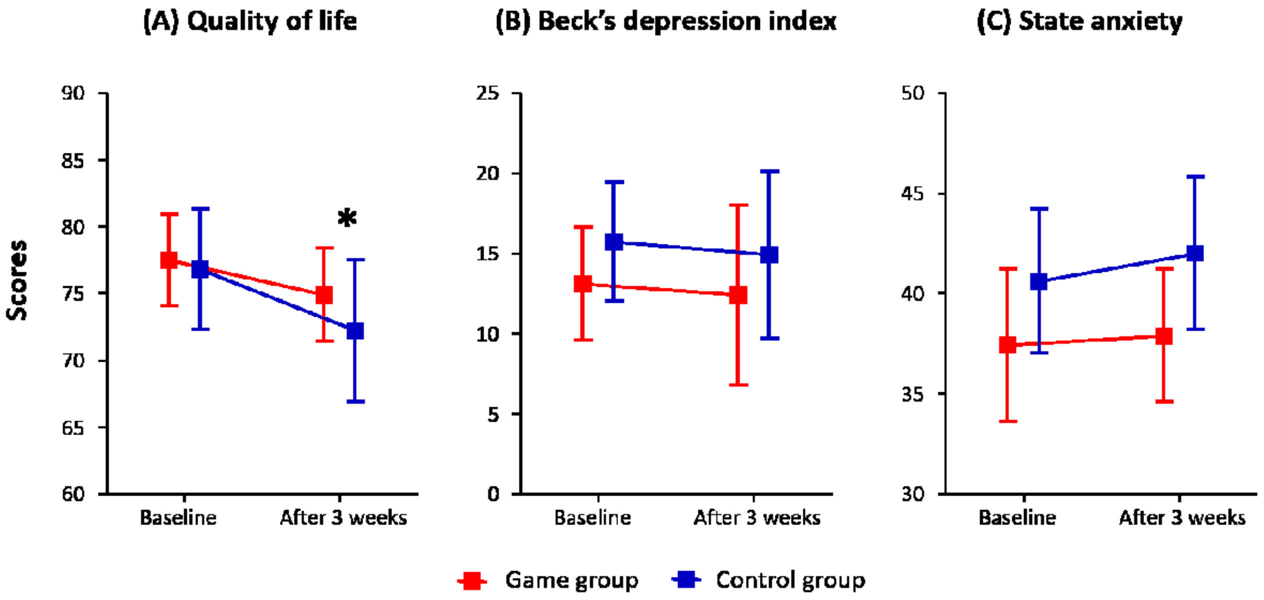
Education time and medication compliance	Game group (n=34)	Control group (n=38)	*P* value
Time spent for game playing or self-education (minutes/day), mean (SD)	22.2 (6.1)	5.5 (4.0)	<.001
Korean Medication Adherence Rating Scale, mean (SD)	7.6 (0.7)	6.5 (0.5)	<.001
Any physical adverse events, n (%)			
Nausea	29 (85)	23 (61)	.02
Fatigue	16 (47)	29 (76)	.02
Decreased appetite	16 (47)	11 (29)	.18
Numbness of hand/foot	0 (0)	22 (58)	.02
Stomatitis	0 (0)	4 (11)	.15
Gastrointestinal (diarrhea or constipation)	7 (21)	9 (24)	.97
Hair loss	0 (0)	10 (26)	.27
Skin rash	0 (0)	0 (0)	N/A[a]
Physical adverse events grade ≥3 of Common Terminology Criteria for Adverse Events, n (%)			
Nausea	5 (15)	0 (0)	.02
Fatigue	1 (3)	12 (32)	.002
Decreased appetite	3 (9)	6 (16)	.59
Numbness of hand and foot	0 (0)	3 (8)	.28
Stomatitis	0 (0)	3 (8)	.28
Gastrointestinal (diarrhea or constipation)	1 (3)	5 (13)	.25
Hair loss	0 (0)	8 (21)	.01
Skin rash	0 (0)	0 (0)	N/A

[a] N/A: not applicable.

Table 3. Comparison of psychological side effects between study groups.

Psychological side effects	Baseline			At 3 weeks			Differences		
	Game, mean (SD)	Control, mean (SD)	*P* value	Game, mean (SD)	Control, mean (SD)	*P* value	Game, mean (SD)	Control, mean (SD)	*P* value
Quality of life	77.5 (3.4)	76.8 (4.5)	.48	74.9 (3.5)	72.2 (5.3)	.01	−2.6 (1.5)	−4.6 (4.4)	.01
Overall quality	2.3 (0.7)	2.3 (0.6)	.74	2.5 (0.7)	2.5 (0.6)	.85	−0.3 (0.3)	0.4 (0.2)	.65
Overall health	1.9 (0.5)	1.7 (0.6)	.14	2.1(0.7)	2.3 (0.6)	.47	0.3 (0.6)	0.6 (0.9)	.08
Physical health	19.1 (2.2)	18.7 (1.8)	.37	20.4 (2.2)	21.1 (1.7)	.16	1.3 (1.0)	2.4 (1.9)	.003
Psychological health	18.1 (1.4)	17.6 (43.1)	.31	18.7 (1.4)	19.3 (4.3)	.43	−2.1 (0.7)	1.7 (2.9)	.02
Social relationships	9.3 (1.3)	9.6 (1.6)	.37	9.4 (1.2)	9.9 (1.9)	.22	0.1 (0.3)	0.2 (2.1)	.67
Environment	23.8 (1.7)	22.9 (3.5)	.16	24.3 (1.9)	24.3 (3.3)	.90	0.4 (1.1)	1.4 (2.5)	.03
Beck's depression index	13.1 (3.5)	12.4 (5.6)	.51	15.7 (3.7)	14.9 (5.2)	.50	2.6 (1.1)	2.6 (1.7)	.99
State anxiety	37.4 (3.8)	37.9 (3.3)	.43	40.6 (3.6)	42.0 (3.8)	.11	3.4 (0.9)	4.1 (3.4)	.21

Figure 4. Psychological adverse events: (A) quality of life, (B) Beck's depression index, and (C) state anxiety. *: P<.05.



Discussion

Principal Findings

The mobile game ILOVEBREAST was developed to help patients with advanced breast cancer learn more about the disease course, properties of medications, and expected adverse drug reactions. Adults users were generally satisfied with the game app. The mobile game-based intervention improved patient compliance, decreased the prevalence of physical side effects, and maintained the patients' QoL compared with the conventional care. The use of the game had no impact on psychological side effects including mood and anxiety.

Improved Drug Compliance

The patients in the study group used approximately 40% of the game contents, and the overall satisfaction was acceptable. Notably, a high proportion of the patients expressed a willingness to play the game again at their next chemotherapy session and to recommend the game to other patients. In addition, education time and drug adherence were significantly higher in the game group than in the control group. The quests for taking medication and reminder alarms in the game may have contributed to improving patients' knowledge of the disease, medication, and adverse events. Previous studies have shown that knowledge about the disease course and adverse drug events is closely linked to patient drug adherence [18]. Since the release of the "Re-Mission" titles, which showed good adherence to treatments and easy access to knowledge, from HopeLab in California, the use of computer games in health education and physical education has been considered to have positive effects [19]. Some video games have also been used as vehicles to transmit health education regarding fire and street safety, and self-management of diabetes and asthma [20,21]. Although most health education games have focused on children, adolescents, and young adults [22-25], this study implies that adults can also utilize and benefit from game-based learning if the contents are specifically designed for them.

Physical Side Effects and Quality of Life

The ILOVEBREAST game presents an avatar that can prevent side effects of numbness and hair loss by purchasing gloves or hats. The users are rewarded for continuous medication use, and they can chat with other players who have the same disease and difficulties. The player can buy food ingredients and learn how to cook and prepare healthy diets. We assume that such activities resulted in decreased prevalence and severity of the physical side effects in the game group. Twitter, as a complementary method, was shown to be effective in increasing knowledge about overall disease course, survivorship, cancer types and biology, and treatment options in patients with breast cancer in a previous study [26]. The Web-based game properties of immersivity, attention-maintaining properties of stories, engaging properties of interactivity, and behavior-change technology may improve health-related behaviors and habits in patients with breast cancer [27]. Several video games have been created to distract people from acute or chronic pain. The “Re-Mission” from HopeLab in California is the most representative mobile game that has these positive effects [19].

The mobile game group in this study showed a higher QoL in various domains, including total health, physical health, psychological health, and environmental areas. Pompeu et al reported improved QoL in 7 patients with Parkinson disease by a game named Kinect Adventures! (Xbox 360, Microsoft Game Studios, United States) [28]. Reichlin et al also reported an improvement in the QoL in 13 patients with localized prostate cancer after playing an interactive game, Time After Time (AMIGO, Gigamic, Piatnik, Playroom Entertainment, Germany) [29].

However, we observed no improvement in the psychological effects in this study. Chen et al reported that pediatric patients with obesity who played educational games for about 10 weeks-2 years improved their psychosocial functions, including depression, self-efficacy, and self-esteem [30]. One possible explanation is the short study duration (ie, 3 weeks) of this research. The age of the participants may be another factor. Unlike children and adolescents, older patients may have a lower propensity to have interest in playing a game.

Study Limitations

The major limitation of this study is the small sample size and the short study period. Future studies are needed to confirm which patients most benefit from the strategy using mobile games and whether the benefits may lead to an improvement in hard endpoints such as mortality or recurrence. Next, adverse events related to chemotherapy were assessed subjectively. Objective assessments such as scales, blood tests, or laboratory tests may better describe the adverse effects of chemotherapy. Also, this study reflects the difficulty when using games for adult patients. Although the patients generally expressed satisfaction, the users experienced only a limited amount of the contents, and more than half had difficulty using the app. We believe this fact provides an important lesson for future developers.

Conclusion

A mobile game, ILOVEBREAST, was helpful in educating adult patients with breast cancer receiving cytotoxic chemotherapy. The game was associated with improved drug compliance, decreased prevalence rates of physical side effects, and better QoL. Patient education with smartphone mobile games can be used as an easy, fun, and effective measure to promote treatment adherence, which may potentially lead to improved survival.

Acknowledgments

This research study was supported by a grant of Nexon 2014 and a grant from the Korea Creative Content Agency, Ministry of Culture, Sports and Tourism (2013040436).

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-EHEALTH checklist (V1.6.1).

[PDF File (Adobe PDF File), 90KB - jmir_v20i10e273_app1.pdf]

References

1. Mumtaz S. Children's enjoyment and perception of computer use in the home and the school. Computers & Education 2001 May 4;36(4):347-362 [FREE Full text] [doi: 10.1016/S0360-1315(01)00023-9]
2. 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]
3. Ryhänen AM, Siekkinen M, Rankinen S, Korvenranta H, Leino-Kilpi H. The effects of Internet or interactive computer-based patient education in the field of breast cancer: a systematic literature review. Patient Educ Couns 2010 Apr;79(1):5-13. [doi: 10.1016/j.pec.2009.08.005] [Medline: 19744817]

4. Oblinger D. The Next Generation of Educational Engagement. JIME 2004 May 21;2004(1):10 [FREE Full text] [doi: 10.5334/2004-8-oblinger]
5. Chia SK, Speers CH, D'yachkova Y, Kang A, Malfair-Taylor S, Barnett J, et al. The impact of new chemotherapeutic and hormone agents on survival in a population-based cohort of women with metastatic breast cancer. Cancer 2007 Sep 01;110(5):973-979 [FREE Full text] [doi: 10.1002/cncr.22867] [Medline: 17647245]
6. Von Ah D, Habermann B, Carpenter JS, Schneider BL. Impact of perceived cognitive impairment in breast cancer survivors. Eur J Oncol Nurs 2013 Apr;17(2):236-241. [doi: 10.1016/j.ejon.2012.06.002] [Medline: 22901546]
7. Hahn T, Wall D, Camitta B, Davies S, Dillon H, Gaynon P, et al. The role of cytotoxic therapy with hematopoietic stem cell transplantation in the therapy of acute lymphoblastic leukemia in children: an evidence-based review. Biol Blood Marrow Transplant 2005 Nov;11(11):823-861 [FREE Full text] [doi: 10.1016/j.bbmt.2005.08.035] [Medline: 16275588]
8. Sedjo RL, Devine S. Predictors of non-adherence to aromatase inhibitors among commercially insured women with breast cancer. Breast Cancer Res Treat 2011 Jan;125(1):191-200. [doi: 10.1007/s10549-010-0952-6] [Medline: 20495864]
9. Osterberg L, Blaschke T. Adherence to medication. N Engl J Med 2005 Aug 4;353(5):487-497. [doi: 10.1056/NEJMra050100] [Medline: 16079372]
10. Bhatia S, Landier W, Hageman L, Kim H, Chen Y, Crews KR, et al. 6MP adherence in a multiracial cohort of children with acute lymphoblastic leukemia: a Children's Oncology Group study. Blood 2014 Oct 09;124(15):2345-2353 [FREE Full text] [doi: 10.1182/blood-2014-01-552166] [Medline: 24829202]
11. First M, Spitzer R, Gibbon M, Williams J. Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I), Clinician Version, Administration Booklet. Arlington: American Psychiatric Publishing, Incorporated; 1997.
12. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR). Arlington: American Psychiatric Publishing, Incorporated; 2000.
13. Jhin-Goo C, Daeyoung R, Chan-Hyung K. The Reliability and Validity of the Korean Version of Medication Adherence Rating Scale. Korean J Psychopharmacol. Korean J Psychopharmacol 2015;26(2):43-49.
14. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry 1961 Jun;4:561-571. [Medline: 13688369]
15. Charles S, Richard G, Robert L, Vagg PR, Gerard J. Manual for the State-Trait Anxiety Inventory. Arlington: Consulting Psychologists Press, Inc; 1983.
16. Min SK, Kim KI, Lee CI, Jung YC, Suh SY, Kim DK. Development of the Korean versions of WHO Quality of Life scale and WHOQOL-BREF. Qual Life Res 2002 Sep;11(6):593-600. [Medline: 12206580]
17. Spielberger C, Gorsuch R, Lushene R, Vagg P, Jacobs G. Manual for the State-Trait Anxiety Inventory. Arlington: Consulting Psychologists Press, Inc; 1983.
18. Zhao S, Zhao H, Wang L, Du S, Qin Y. Education is critical for medication adherence in patients with coronary heart disease. Acta Cardiol 2015 Apr;70(2):197-204. [doi: 10.2143/AC.70.2.3073511] [Medline: 26148380]
19. Kato PM, Cole SW, Bradlyn AS, Pollock BH. A video game improves behavioral outcomes in adolescents and young adults with cancer: a randomized trial. Pediatrics 2008 Aug;122(2):e305-e317 [FREE Full text] [doi: 10.1542/peds.2007-3134] [Medline: 18676516]
20. Brown SJ, Lieberman DA, Germeny BA, Fan YC, Wilson DM, Pasta DJ. Educational video game for juvenile diabetes: results of a controlled trial. Med Inform (Lond) 1997;22(1):77-89. [Medline: 9183781]
21. Bartholomew LK, Gold RS, Parcel GS, Czyzewski DI, Sockrider MM, Fernandez M, et al. Watch, Discover, Think, and Act: evaluation of computer-assisted instruction to improve asthma self-management in inner-city children. Patient Educ Couns 2000 Feb;39(2-3):269-280. [Medline: 11040726]
22. Capel S, Whitehead M. The use of information and communication technology in physical education. In: Learning to Teach Physical Education in the Secondary School: A Companion to School Experience. Abingdon-on-Thames: Taylor & Francis, Inc; 2010.
23. Casazza K, Ciccazzo M. Improving the dietary patterns of adolescents using a computer-based approach. J Sch Health 2006 Feb;76(2):43-46. [doi: 10.1111/j.1746-1561.2006.00067.x] [Medline: 16466465]
24. Coles CD, Strickland DC, Padgett L, Bellmoff L. Games that "work": using computer games to teach alcohol-affected children about fire and street safety. Res Dev Disabil 2007;28(5):518-530. [doi: 10.1016/j.ridd.2006.07.001] [Medline: 16965894]
25. Neuhauser L, Kreps GL. Rethinking Communication in the E-health Era. J Health Psychol 2003 Jan;8(1):7-23. [doi: 10.1177/1359105303008001426] [Medline: 22113897]
26. Attai DJ, Cowher MS, Al-Hamadani M, Schoger JM, Staley AC, Landercasper J. Twitter Social Media is an Effective Tool for Breast Cancer Patient Education and Support: Patient-Reported Outcomes by Survey. J Med Internet Res 2015;17(7):e188 [FREE Full text] [doi: 10.2196/jmir.4721] [Medline: 26228234]
27. Baranowski T, Buday R, Thompson DI, Baranowski J. Playing for real: video games and stories for health-related behavior change. Am J Prev Med 2008 Jan;34(1):74-82 [FREE Full text] [doi: 10.1016/j.amepre.2007.09.027] [Medline: 18083454]
28. Pompeu JE, Arduini LA, Botelho AR, Fonseca MBF, Pompeu SMAA, Torriani-Pasin C, et al. Feasibility, safety and outcomes of playing Kinect Adventures!™ for people with Parkinson's disease: a pilot study. Physiotherapy 2014 Jun;100(2):162-168. [doi: 10.1016/j.physio.2013.10.003] [Medline: 24703891]

29. Reichlin L, Mani N, McArthur K, Harris AM, Rajan N, Dacso CC. Assessing the acceptability and usability of an interactive serious game in aiding treatment decisions for patients with localized prostate cancer. J Med Internet Res 2011 Jan 12;13(1):e4 [FREE Full text] [doi: 10.2196/jmir.1519] [Medline: 21239374]
30. Chen J, Wilkosz ME. Efficacy of technology-based interventions for obesity prevention in adolescents: a systematic review. Adolesc Health Med Ther 2014;5:159-170 [FREE Full text] [doi: 10.2147/AHMT.S39969] [Medline: 25177158]

Abbreviations

BDI: Beck Depression Inventory
K-MARS: Korean version of the Medication Adherence Rating Scale
QoL: quality of life

Edited by G Eysenbach; submitted 04.12.17; peer-reviewed by L Lam, JL Chen, M Shiyko; comments to author 18.03.18; revised version received 23.06.18; accepted 10.07.18; published 29.10.18.

Please cite as:
Kim HJ, Kim SM, Shin H, Jang JS, Kim YI, Han DH
A Mobile Game for Patients With Breast Cancer for Chemotherapy Self-Management and Quality-of-Life Improvement: Randomized Controlled Trial
J Med Internet Res 2018;20(10):e273
URL: http://www.jmir.org/2018/10/e273/
doi:10.2196/jmir.9559
PMID:30578205

Original Paper

Proposing a Transactional Model of eHealth Literacy: Concept Analysis

Samantha R Paige[1,2], PhD, MPH; Michael Stellefson[3], PhD; Janice L Krieger[1], PhD; Charkarra Anderson-Lewis[4], PhD, MPH; JeeWon Cheong[2], PhD; Christine Stopka[2], PhD, AT Ret, CSC-Ret, CAPE

[1]STEM Translational Communication Center, College of Journalism and Communications, University of Florida, Gainesville, FL, United States

[2]Department of Health Education & Behavior, University of Florida, Gainesville, FL, United States

[3]Department of Health Education and Promotion, East Carolina University, Greenville, NC, United States

[4]Department of Public Health, University of Southern Mississippi, Hattiesburg, MS, United States

Corresponding Author:
Samantha R Paige, PhD, MPH
STEM Translational Communication Center
College of Journalism and Communications
University of Florida
PO Box 118400
Gainesville, FL, 32611
United States
Phone: 1 352 294 0421
Email: paigesr190@ufl.edu

Abstract

Background: Electronic health (eHealth) literacy was conceptualized in 2006 as the ability of internet users to locate, evaluate, and act upon web-based health information. Now, advances in eHealth technology have cultivated transactional opportunities for patients to access, share, and monitor health information. However, empirical evidence shows that existing models and measures of eHealth literacy have limited theoretical underpinnings that reflect the transactional capabilities of eHealth. This paper describes a conceptual model based on the Transactional Model of Communication (TMC), in which eHealth literacy is described as an intrapersonal skillset hypothesized as being dynamic; reciprocal; and shaped by social, relational, and cultural contexts.

Objective: The objective of our study was to systematically examine eHealth literacy definitions, models, and measures to propose a refined conceptual and operational definition based on the TMC.

Methods: Walker and Avant's concept analysis method was used to guide the systematic review of eHealth literacy definitions (n=10), rating scales (n=6), models (n=4), and peer-reviewed model applications (n=16). Subsequent cluster analyses showed salient themes across definitions. Dimensions, antecedents, and consequences reflected in models and measures were extracted and deductively analyzed based on codes consistent with the TMC.

Results: Systematic review evidence revealed incongruity between operational eHealth literacy included in definitions compared with literacies included within models and measures. Theoretical underpinnings of eHealth literacy also remain dismal. Despite the transactional capabilities of eHealth, the role of "communication" in eHealth literacy remains underdeveloped and does not account for physical and cognitive processing abilities necessary for multiway transactions.

Conclusions: The Transactional Model of eHealth Literacy and a corresponding definition are proposed. In this novel model, eHealth literacy comprises a hierarchical intrapersonal skillset that mediates the reciprocal effect of contextual factors (ie, user oriented and task oriented) on patient engagement in health care. More specifically, the intrapersonal skillset counteracts the negative effect of "noise" (or impediments) produced by social and relational contexts. Cutting across health and technology literacies, the intrapersonal skillset of eHealth literacy is operationalized through four literacies that correspond with discrete operative skills: (1) functional (ie, locate and understand); (2) communicative (ie, exchange); (3) critical (ie, evaluate); and (4) translational (ie, apply).

(J Med Internet Res 2018;20(10):e10175) doi:10.2196/10175

KEYWORDS

eHealth literacy; Transactional Model of Communication; interpersonal communication; social media; mobile phone

Introduction

Electronic health (eHealth) is increasingly being ingrained within the health care system and patient engagement experience. eHealth facilitates productive collaborations among informed patients, proactive health care professionals, and responsive health care systems to coordinate care for positive health outcomes [1,2]. Alongside the evolution of eHealth, patients can now interact with health information available on the internet and either synchronously or asynchronously exchange ideas, thoughts, and health-related data and media with other users through multimedia on computer-mediated platforms (eg, health information portals, personal health records, telemedicine apps, Web-based support groups or forums) [1,2]. Shaw et al [2] identified 3 overlapping domains of eHealth, including users' interaction with technology, interaction with other users through mediated platforms, and use of information gained from these interactions to advance their health and well-being. As such, a core aspect of eHealth includes not only the use of technology but also the computer-mediated transaction of information among its users.

The Transactional Model of Communication (TMC) [3] posits that communication between two or more entities is dynamic, process oriented, and adapted or appropriated according to the context of the transaction. This context is shaped by the channel of communication (eg, telephone, email, letter), the source of communicators (eg, interpersonal, impersonal), language (eg, native, second), and the type of message (eg, mode of transmission, whether image, video, text, or other). Social, relational, and cultural contexts also drive the transactional process of communication. In the TMC, entities are not assigned roles as message "senders" or "receivers"; rather, their roles are interdependent, meaning that they are simultaneously message senders and receivers or simply communicators. Any person within a social situation is a communicator, whether his or her interaction is synchronous or asynchronous, verbal or nonverbal, and intentional or unintentional. In this model, communication extends beyond a simplistic view of message creation; the model views processing information as a vehicle for community and personal identity construction and impression management within the transactional context (eg, source, channel, message, language) [3-5]. The TMC functions under the assumption that interpersonal communication exists within a fluid state and that the transaction among communicators is constantly changing and mutually influenced.

The TMC can be extended to interpersonal computer-mediated communication (I-CMC). I-CMC occurs remotely with technology (eg, desktop computer, smartphone, tablet, laptop) through diverse message channels (eg, text, video, image) and sources (eg, personal friends and family, impersonal provider, peer) [6]. In the social era of eHealth, or Web 2.0, where two-way transactions occur among users, the device and channel drive the type and amount of information transmitted by diverse sources [7-9]. I-CMC notoriously fosters ambiguous communication because traditional in-person social and contextual cues that assist people in understanding the pragmatic meaning of messages are less salient across computer-mediated platforms [5,7]. With such cues filtered out, I-CMC can disrupt the accurate and smooth transmission of multimedia messages among communicators using various channels [5]. Similar to in-person communication, there are factors beyond contextual and social cues in the TMC that can exacerbate the ambiguity of message transmission via I-CMC.

Noise-inducing factors interfere with information transmission and accessibility among communicators, ultimately hindering their ability to access, understand, and transmit meaning to one another [10]. Noise-induced factors can be categorized as physical (ie, external factors), psychological (ie, mental and emotional belief-systems), physiological (ie, physical conditions, including auditory and verbal limitations, and medication effects), and semantic (ie, systems of meaning that do not correspond) [6]. With regard to I-CMC in the context of eHealth, these noise-inducing factors include technological usability challenges, stress or worry related to a recent disease diagnosis, exposure to scientific medical jargon, and physical limitations due to a health condition, just to name a few. Generally, "noise" can be compared with barriers or impediments widely published in the literature to describe hindrances to successful eHealth adoption and use [11-13]. In the TMC from the perspective of health-related I-CMC, however, barriers are operationalized beyond functional technological impediments; rather, they act as personal, relational, social, and cultural factors that hinder the process of communication [3,6]. As such, the high volume and constant flow of health information created and shared on the internet, coupled with the regular presence of noise-inducing factors, has the potential to attenuate users' capacity to effectively and appropriately engage in the transmission of health-related communication. An essential aspect of successful transactional communication within computer-mediated contexts is the capacity to counteract the negative effects of noise.

To understand patients' capacity to successfully use and benefit from eHealth, the concept of eHealth literacy was initially coined in 2006. eHealth literacy was defined as the ability to locate, evaluate, understand, and act upon health information from electronic sources [14,15]. Despite widespread use of this definition over the past decade, researchers have argued that this seminal construct and its corresponding eHealth Literacy Scale (eHEALS) are outdated because neither considers the evolving dynamic and social nature of eHealth [2,16-18]. In an attempt to synthesize eHealth literacy research and recommendations for its conceptual advancement, Griebel et al [19] posited that new eHealth literacy concepts do not build upon the assumptions and structure of existing models; rather, these models function in isolation and do not emanate from the existing literature. Moreover, empirical evidence shows that eHealth literacy definitions and models have insufficient theoretical underpinnings, which inhibits eHealth literacy researchers from developing an updated definition, model, and corresponding measure that reflects the social context of eHealth [19]. Together, these limitations perpetuate challenges in advancing our understanding of eHealth literacy in the social era of eHealth, specifically regarding its valid operationalization and measurement.

The purpose of this study is to propose a theoretical blueprint for defining and conceptualizing eHealth literacy in the transactional era of eHealth. Per the fundamental assumptions of the TMC, this study operationalizes eHealth literacy as an intrapersonal skillset grounded in counteracting the effect of noise during transactional interactions across computer-mediated platforms. In the context of I-CMC, the TMC is appropriate to form the basis of our proposed model because we aim to describe the communicative element of eHealth literacy and understand how underlying eHealth operational skills function in the larger context of computer-mediated transactions. In this study, we applied a concept analysis method, which is a rigorous method in which empirical literature is systematically surveyed to refine the operationalization of a construct [20,21]. The findings of this empirical review generate an operational definition and eHealth literacy model based on the TMC.

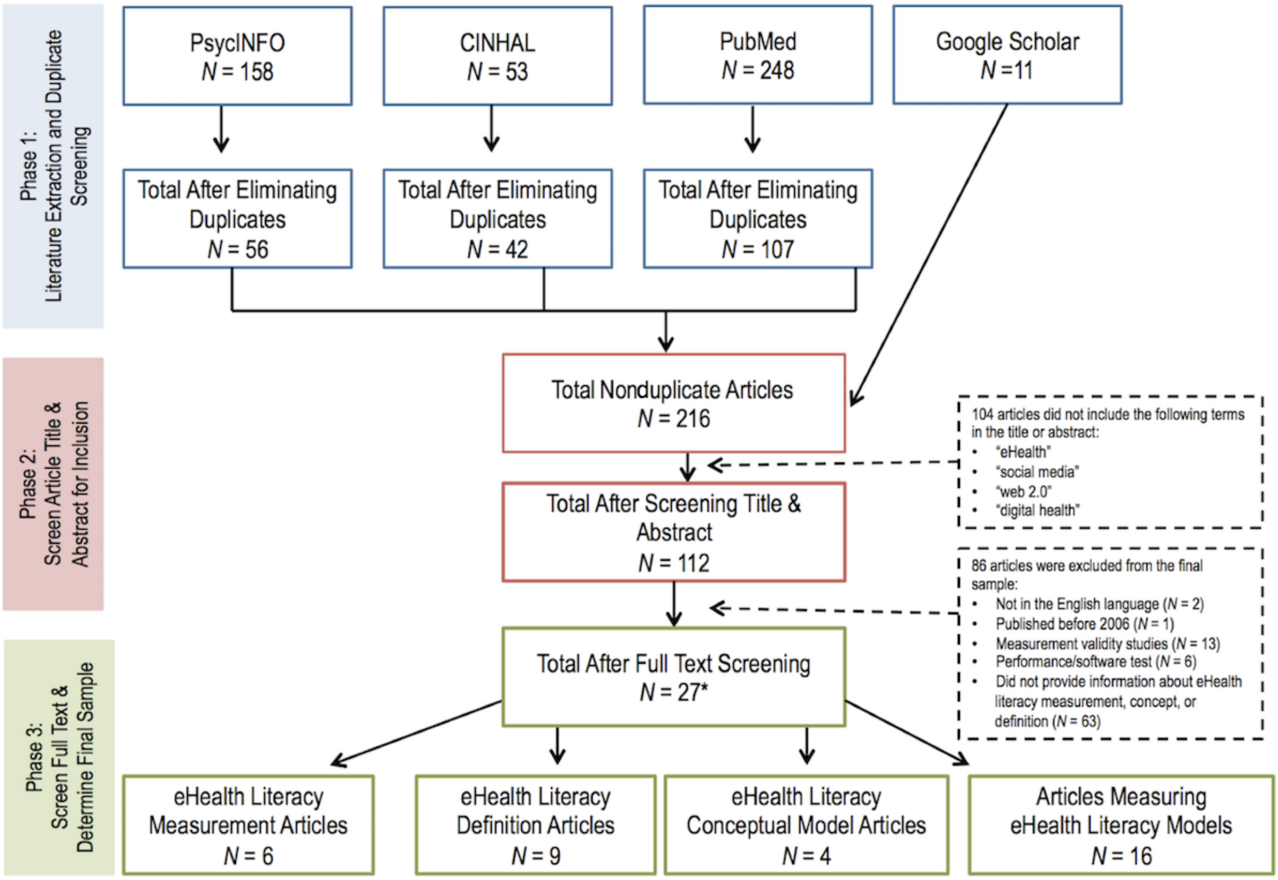
Methods

Sample and Procedures: Concept Analysis

A series of keywords were combined with the Boolean operator ("AND") and entered into 3 electronic databases (ie, PubMed, CINAHL, and PsycINFO). In each search query, a combination of 3 terms was entered to reflect the following: (1) *purpose* (ie, "concept," "model," "definition," "framework," "theory," "measure," "instrument," "scale," and "survey"); (2) *context* (ie, "eHealth," "social media," "Web 2.0," "social network," and "digital health"); and (3) *ability* (ie, "skill" and "literacy"). The same queries were conducted in Google Scholar to identify gray (or unpublished) literature. The final sample consisted of articles that (1) were published between 2006 and 2017 (2006 is the year that the seminal eHealth literacy definition, model, and measure were published); (2) were in the English language; (3) included the terms "eHealth," "social," "media," "Web 2.0," or "digital health" in the title or abstract; and (4) presented information on the concept, definition, or measurement of skills related to social media, digital health, and electronic health record use. Figure 1 presents literature review extraction procedures. Twenty-seven unique articles met eligibility criteria. Some articles included both definitions and models or measures; therefore, the asterisk indicates that Phase 3 N values exceed the total sample of 27.

Walker and Avant's [21] concept analysis methodology was used to guide the data extraction and analysis procedures. Literature presenting definitions, antecedents, consequences, and attributes (ie, dimensions) of eHealth literacy was extracted. Articles that presented explicit definitions and conceptual models of eHealth literacy were considered. Moreover, the original sources of eHealth literacy empirical referents (ie, measurement instruments) were included in the final sample. Peer-reviewed empirical articles that included at least one of the models reviewed in the analyses were perused to identify information about antecedents and consequences of eHealth literacy.

Figure 1. Literature extraction procedures. The asterisk indicates that Phase 3 N values exceed the total sample of 27 as some articles included both definitions and models or measures.



Data Analysis

An inductive analysis of eHealth literacy definitions was performed to identify thematic clusters from peer-reviewed and gray literature. Consistent with a separate concept analysis of health literacy [22], an inductive analysis of eHealth literacy definitions identified thematic clusters by *competence*, *contextual factors*, *action (operational behaviors or skills)*, *object of interest*, and *objective*. The attributes of eHealth literacy conceptual models and measures were extracted and entered into a descriptive table where congruent components were identified; furthermore, antecedents (independent variables in analyses) and consequences (dependent variables) were extracted.

Results

Concept Analysis: eHealth Literacy Definitions, Models, and Measures

Existing Definitions of eHealth Literacy

Table 1 presents 10 eHealth literacy definitions published between 2006 and 2017. Seminal definitions solely focused on intrapersonal skills to access and use health information obtained from electronic sources [14,23]. Interactions between individual and technological factors became more salient in later definitions of eHealth literacy [17,24]. Chan and Kaufman [25], for example, posited that eHealth literacy is not solely dependent on cognitive processing; rather, it is influenced through interactions between cognition and technology. More recent definitions of eHealth literacy stated that eHealth skills function within the context of social, individual, and technological factors [16,19,26]. The interaction between diverse contextual factors and technological constraints influences eHealth skills and the ultimate capacity to improve health and wellness. Although implied in all definitions, one definition explicitly stated that eHealth literacy comprises a “hybrid of two other concepts,” including health literacy and technology literacies [27].

Textbox 1 presents the definitions of eHealth literacy into clustered themes. *Competence* is characterized as a set of skills and knowledge, predominantly referred to as “the ability.” *Influential Factors* that determine the said ability are characterized as the interplay between contextual factors (ie, individual and social) coupled with situational factors (ie, the type of health problem and type of technology).

Table 1. Ten definitions of eHealth literacy from the research literature (2006-2017).

Article	Author	Year	Definition
1	Norman & Skinner [14]	2006	*The ability to seek out, find, understand and appraise, integrate, and apply what is gained in electronic environments toward solving a health problem.* [pg 2]
2	Bodie & Dutta [23]	2008	*…not just the ability to use the internet to find answers to health-related questions; it also entails the ability to understand the information found, evaluate the veracity of the information, discern the quality of different websites, and use the quality information to make informed decisions about health.* [pg 193]
3	Chan & Kaufman [25]	2011	*A set of skills and knowledge that are essential for productive interactions with technology-based health tools, such as proficiency in information retrieval strategies, and communicating health concepts effectively.* [pg 2]
4	Norman [17]	2011	*A foundational skill set that underpins the use of information and communication technologies for health.* [pg 1]
5	Neter & Brainin [24]	2012	*The ability of people to use emerging information and communication technologies to improve or enable health and health care.* [pg 1]
6	Paek & Hove [27]	2012	*…a hybrid of two other concepts, eHealth and health literacy, [in which] skills must be appropriate for the informational text people need to understand in their efforts to treat various health concerns.* [pg 728]
7	Werts & Hutton-Rogers [29]	2013	*The ability to gather and appropriately process health information retrieved online.* [pg 115]
8	Gilstad [26]	2014	*The ability to identify and define a health problem, to communicate, seek, understand, appraise, and apply eHealth information and welfare technologies*[a] *in the cultural, social and situational frame and to use the knowledge critically in order to solve the health problem.* [pg 69]
9	Bautista [16]	2015	*The interplay of individual and social factors in the use of digital technologies to search, acquire, comprehend, appraise, communicate, and apply health information in all contexts of health care with the goal of maintaining or improve the quality of life throughout the lifespan.* [pg 43]
10	Griebel et al [19]	2017	*…a dynamic and context-specific set of individual and social factors, as well as consideration of technological constraints in the use of digital technologies to search, acquire, comprehend, appraise, communicate, apply, and create health information in all contexts of health care with the goal of maintaining or improving the quality of life throughout the lifespan.* [pg 10]

[a] Welfare technologies: “strengthen a users’ independence, safety, control of surroundings, independent living and social activities, independent of age and disabilities” (pg 344) [28].

Textbox 1. Five clusters of eHealth literacy definitions from the literature. Numerals in brackets correspond with article numbers in Table 1.

Competence

- Ability (1,2,5,7,8)
- Skills (6)
- Set of skills & knowledge (3)
- Foundational skillset (4)

Contextual factors

- Cultural, social, and situational frame (8)
- Interplay between social and individual factors in using technology (9)
- Dynamic and context-specific individual and social factors and technological constraints (10)
- To use information and communication technologies (4,5)

Action

- To locate
    - Find (1,2)
    - Seek (1,8)
    - Search (9,10)
    - Retrieve (3)
    - Gather (7)
    - Acquire (9,10)
- To understand
    - Comprehend (9,10)
    - Understand (1,2,6,8)
    - Process (7)
- To evaluate
    - Appraise (1,8,9,10)
    - Evaluate the veracity (2)
    - Discern the quality (2)
- To communicate (3,8,9,10)
- To create (10)
- To translate
    - Integrate (1)
    - Apply (1,8,9,10)
    - Use knowledge (8)

Object of interest

- Knowledge (1,8)
- Information
    - General (2,4)
    - Health (9,10)
    - Quality (2)
    - Emerging (5)
    - eHealth (8)

  - Online (7)
  - Text (6)

Objective

- For health (4)
- To address or solve a health problem (1,8)
- To make informed decisions about health (2)
- To improve or enable health and health care (5)
- To treat various health concerns (6)
- To maintain or improve the quality of life throughout the lifespan (9,10)

Actions are thematically clustered according to operational skills, including the capacity to locate, understand, evaluate, exchange, and apply or translate health information. The *Object of Interest*, or the purpose of performing the actions, includes obtaining knowledge from high-quality Web-based health information. Finally, the *Objective* of obtaining the object of interest is generally for the purposes of health enhancement or to maintain or improve the health-related quality of life throughout the lifespan.

Dimensions of eHealth Literacy

Since 2006, 4 models and 6 measurement instruments of eHealth literacy have been published; their purposes, guiding theoretical frameworks (if applicable), and dimensions used in describing the concept and measurement have been presented in Multimedia Appendix 1.

Conceptual Models (N=4)

Norman and Skinner's [14] Lily Model posits that eHealth literacy is comprised of analytic skills central to the Web-based health information seeking experience, as well as context-specific skills that vary according to situation. In the Lily Model, health science and computer literacies are denoted as context-specific skills, whereas information, traditional literacy and numeracy, and media literacies are analytic-specific. As such, a high degree of eHealth literacy exists where individuals use context- and analytic-specific skills in concert, and it allows an individual to successfully achieve an eHealth goal.

Gilstad [26] adapted Norman and Skinner's [14] Lily Model to describe how eHealth literacy mediates the effect of diverse contextual factors (or literacies) on productive patient-provider communication. This model posits that contextual "literacies" (ie, propositional, cultural, social, propositional, and procedural) coupled with situational factors (ie, type of health question and type of eHealth technology) directly impact context- and analytic-specific aspects of eHealth literacy as posed in Norman and Skinner's Lily Model. In Gilstad's model, the outcome associated with eHealth literacy is communicative expertise (ie, the capacity to discuss a personal or family concern with an offline health care provider).

Unlike Norman and Skinner [14] and Gilstad [26], Bautista [16] developed a model to posit that eHealth literacy is a process-oriented concept. In this model, Bautista states that eHealth literacy is reciprocal, meaning that this construct affects and is affected by diverse contextual and ecological factors. As such, Bautista defines eHealth literacy as comprising intrapersonal actions (ie, search, acquire, comprehend, appraise, communicate, and apply), the type of digital technology selected (ie, PC and mobile devices), the Web-based environment in which the search occurs (ie, social media vs traditional website), as well as the goal of using eHealth technologies (ie, maintenance and treatment) in particular health care contexts (ie, promotion, prevention, curative, and rehabilitation) across the lifespan.

Kayser et al [30] applied an informatics approach to conceptualize eHealth literacy through a multidisciplinary lens. This model applies a user-task-context matrix, grounded in health and digital literacy. The matrix is comprised of 7 elements from 3 domains (ie, user, task, and user-task). The model functions under the assumption that eHealth literacy is the degree of harmony between health care consumers' needs and skills, as well as the capacity of the technology to meet those needs and foster those skills within the greater health care context.

Conceptual models have attempted to extend Norman and Skinner's [14] Lily Model to depict how contextual factors influence individual eHealth skills. Gilstad [26] identified a number of situational, technological, and cultural factors that can influence the intrapersonal literacies outlined in the Lily Model. Kayser et al [30], who did not consider the Lily Model in their conceptualization of eHealth literacy, roughly defined influential contextual factors as user- and task-domains, positing that eHealth skills are dependent on both the situation and person. Only Bautista's [16] model depicts operational skills that include "communication" as a central skill. Bautista's model also depicts eHealth literacy as intrapersonal skills that have reciprocal relationships with contextual factors, the type of technology, personal factors (ie, age), and the purpose of the eHealth experience.

Synthesized together, these 4 models suggest that eHealth literacy is an intrapersonal skillset shaped by diverse contextual factors influencing the user and the situation prompting the eHealth interaction.

Measurement Instruments (N=6)

Multimedia Appendix 1 also presents the results of 6 self-administered eHealth literacy rating scales, including their purposes, guiding frameworks, and dimensions. Norman and Skinner's [15] eHEALS was the seminal eHealth literacy measurement instrument, developed as a unidimensional scale grounded in self-efficacy to reflect the Lily Model [14]. Chew and Yuqian [31] refute that the eHEALS reflects dimensions of the Lily Model. Their measure identified items from the Health Information National Trends Survey to address each of the literacies within the Lily Model. Unfortunately, insufficient evidence was reported to support the psychometric properties of their generic eHealth literacy measure.

Other measurement instruments assessed motivation or readiness to use eHealth. Through formative focus groups with older adults with chronic diseases, Koopman et al [32] derived 8 dimensions of eHealth motivation, including (but not limited to) the need for health information, preferred mode and channel of eHealth interaction, and electronic privacy, but these dimensions were not based on an existing conceptualization of eHealth literacy. Bhalla et al [33] assessed eHealth readiness by conducting formative research with eHealth consumers to identify themes that corresponded with the constructs of Social Cognitive Theory, specifically self-regulation to use eHealth. Despite being grounded in a health behavior theory, the dimensions included within this measure do not reflect the central components of a proposed or existing eHealth literacy model.

Most recently, measurement instruments have been developed to account for the social features of eHealth. Seçkin et al [34] identified 3 important concepts from a systematic literature review of health literacy that appeared to be central to eHealth literacy, including cognitive literacy (trust), interactional literacy (communication with offline health care providers), and behavioral literacy (apply learned health behaviors). Additionally, van der Vaart and Drossaert [35] developed an instrument to measure digital health literacy among rheumatic patients. The items on this scale capture dimensions about the capacity to use technology, navigate Web-based health information, create text messages for other users, and take precautions to protect the privacy of themselves and other users. These most recent measurement instruments begin to consider the operational skills related to eHealth proficiency; however, these measures are derived from formative research with limited application to eHealth literacy definitions or conceptual models.

Antecedents and Consequences of eHealth Literacy

Functional (or basic), health, and technology literacies are fundamental to eHealth skills [14,24,29,36,37]. Antecedents that influence eHealth skills include personal, relational, knowledge, and technological determinants. *Personal Determinants* influencing eHealth literacy include income and education [15,18,24,36-39], race or ethnicity [39], gender [40], age [18,24,40-42], marital status [40], and health status [41]. *Relational Determinants* include social influences or norms and alleviated linguistic and cultural barriers to health information [27,29]. *Knowledge Determinants* include the type and amount of health information preferred and the amount of pre-existing knowledge about the health concern [14,24,30,40]. *Technological Determinants* include motivation to use technology for health [15,23,24,27,36,42], access to technological devices [26,43], the type and number of technologies used to access health information [15,18], frequency of using eHealth [24,27,41,44], and preference for using eHealth to help address a particular concern [37].

The consequences of eHealth literacy primarily comprise intrapersonal factors, which have a residual effect across socioecological contexts. The most salient intrapersonal consequence includes a change in the degree of patient engagement [14,29,45]. People with a high degree of eHealth literacy report greater health care access [42]; better health-related outcomes [29,36,43]; and participation in proactive health behaviors offline, including self-management behaviors [24], patient-provider communication [24], and cancer screenings [40]. Consistent with the central tenants of eHealth literacy, a greater degree of confidence in eHealth skills was associated with higher self-reported comprehension [46], critical evaluation [44,45], and trust in Web-based health information from diverse sources and channels [26,39]. Positive self-reported eHealth skills predict motivation to continue using eHealth [23,26], particularly because it is perceived as a useful tool for supplementing health care [44].

Transactional Model of eHealth Literacy

This synthesis and review of eHealth literacy definitions, models, and measures posits that the multidimensional construct is a reciprocal intrapersonal skillset influenced by the interaction between user- and task-oriented factors, which drive patient engagement, empowerment, and informed decision making. This is consistent with the theoretical underpinnings of the TMC [3]. According to the assumptions of the TMC [3], information transaction is dependent on the interaction between a series of contextual factors. In synthesizing the antecedents of transactions based upon eHealth literacy literature and the TMC, the contextual factors that influence eHealth literacy can be categorized as task-oriented features (ie, message type, source, channel, and language) and user-oriented features (ie, personal, relational, knowledgeable, and technological). The interactions between task- and user-oriented factors produce variable degrees of physical, semantic, psychological, and physiological noise [3,6,10]. The effects of the noise can either hinder or facilitate successful transaction of Web-based health information. The intrapersonal skillset of eHealth literacy will be integral for a user to benefit from the eHealth experience. Per eHealth literature, it is probable to hypothesize that the eHealth experience will inform the perceived affordance of eHealth in the future. This reciprocal feature of our model further captures the transactional and continuous elements of eHealth.

Existing definitions and models do not capture the transactional nature of eHealth literacy, specifically regarding information exchange, knowledge application, and message generation when communicating with other users. The most recent eHealth literacy frameworks depict eHealth literacy according to the intersection of user attributes, perceived motivation or control, and experiences using eHealth [30,47]. This framework posits that the ability of an individual to actively engage with digital services is central to the eHealth experience. The operationalization of these abilities, however, appears limited to functional skills related to eHealth or to comfort in handling technologies to learn about health information or enter health-related data [30,46,47]. To build upon these previous frameworks, our proposed definition and model of eHealth literacy considers technology a tool that is used as a vehicle to help users access and exchange health information that can be critically analyzed and translated across socioecological facets of health care. We extend beyond these basic functional behaviors related to technology readiness and engagement and rather consider eHealth literacy as a hierarchical skillset that allows people to not only use technology but also engage with others via technology to participate in dynamic health information seeking and transactional exchanges across computer-mediated platforms. The contribution of our proposed definition and model lies in its ability to position existing eHealth literacy literature through a translational health communication lens. We aim to theoretically capture the transactional nature of eHealth, specifically regarding both technology use and social engagement. The theorized definition and model are proposed as follows.

Proposed Definition

The following definition of *eHealth literacy* is proposed:

> *The ability to locate, understand, exchange, and evaluate health information from online environments in the presence of dynamic contextual factors and to apply the knowledge gained across ecological levels for the purposes of maintaining or improving health.*

This definition builds upon previous eHealth literacy definitions and extends the concept to the context of the TMC. First, operational skills comprise and correspond with the central aspects of eHealth [2]: (1) interaction with technology (ie, *locate, understand*); (2) interaction with other users through mediated platforms (ie, *exchange*); and (3) assessment (ie, *evaluate*) and action (ie, *apply*) for health advancement. Second, the proposed definition acknowledges that eHealth literacy is highly contextual as it varies according to the interplay between user- and task-oriented factors. Third, this definition highlights the affordances of technology, which assist lay end users to access and exchange health information using electronic tools. Based on the synthesis of the literature, proficiency in using a technology or a Web-based environment does not solely determine one's eHealth literacy; rather, it is the capacity of the users to achieve their intended eHealth goals when encountering noise that challenges the successful use of technology and transaction of health information. Finally, consistent with interpersonal communication literature and the TMC [3,5,6], this definition clarifies that "communication" in eHealth literacy is the ability to construct relationships and identities with other Web-based users through health information exchange.

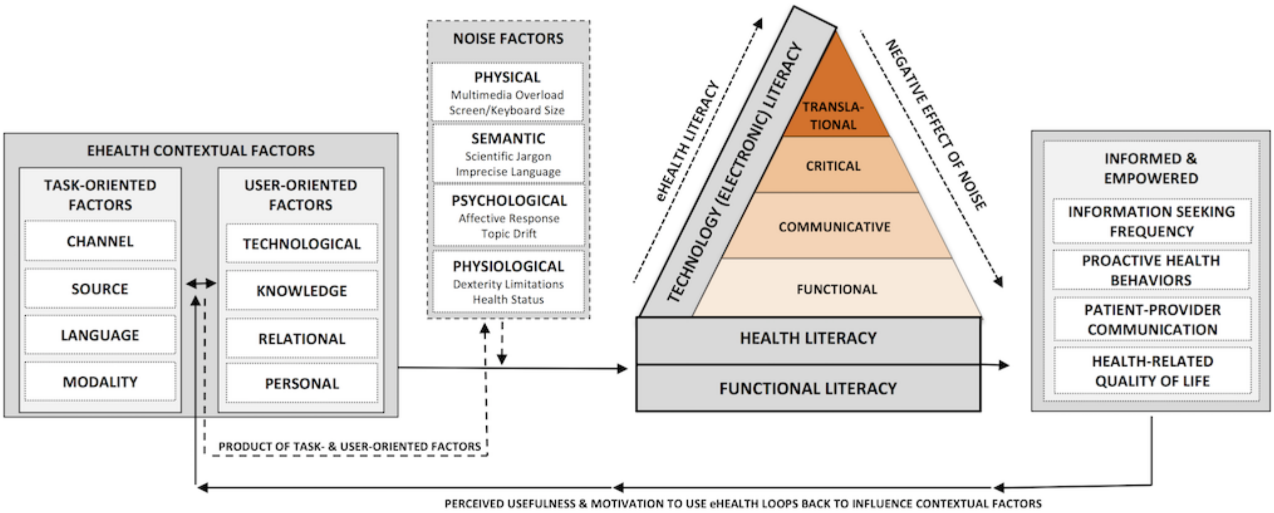
Through this updated definition of eHealth literacy, we extend beyond functional behaviors related to technology readiness and engagement. We consider eHealth literacy as a hierarchical skillset that not only allows users to engage with technology but also engages other users via technology for dynamic health information seeking and transactional exchanges across computer-mediated platforms. The contribution of our proposed definition lies in its ability to position current eHealth literacy literature through a translational health communication lens.

Proposed Model

Consistent with the definition above, Figure 2 presents the Transactional Model of eHealth Literacy (TMeHL), which is derived from a systematic review of the literature (ie, concept analysis) and is theoretically based on the TMC. This model does not specify encoders (sender) and decoders (receiver); rather, it treats the communication transaction as a continuous process that is constantly modified according to diverse eHealth contextual factors and prior eHealth experiences. The TMeHL consists of 3 assumptions: (1) task-oriented and user-oriented factors interact to produce physical, semantic, psychological, and physiological noise during the transaction process; (2) eHealth literacy, a multidimensional and hierarchical intrapersonal skillset, counteracts the effect of noise on the transaction; and (3) patient engagement is reciprocal and influences future interactions between eHealth contextual factors and their effect on eHealth literacy. Although the primary consequence associated with eHealth literacy is being an informed and engaged patient across diverse socioecological contexts, there is no "end goal" of eHealth literacy. The capacity of an informed and engaged patient to apply knowledge gained from an eHealth transaction across diverse socioecological factors (ie, trust in eHealth, productive patient-provider communication, greater eHealth use and perceptions of its usefulness, and positive health-related quality of life) will ultimately inform patients' future eHealth motivation and perceived usefulness for addressing a particular health concern. In turn, these experiences are hypothesized to inform future experiences and perceived affordances of eHealth by shaping task- and user-oriented factors that drive future noise production and eHealth skills.

Consistent with prior eHealth literacy models [26,30], a series of task-oriented and user-oriented factors comprise the eHealth context. However, these factors do not function in isolation, and they extend beyond the ability to interact and use technology; rather, these factors interact with one another to shape the transactional process of eHealth experience, including eHealth intrapersonal skills. *Task-Oriented Factors* include the channel in which the transaction occurs (eg, social media, electronic health record, email), the source or identity of the communicators (eg, peer, friend, family member, health care provider), the language used to communicate (eg, native, second), and the modality of the message (eg, image, text, video). *User-Oriented Factors*, however, comprise factors that are central to the user, rather than to the situation or task. These factors include personal demographic information, including education, gender, and age. Relational support is described as the amount of support or perceived social norm in using eHealth.

Figure 2. The Transactional Model of eHealth Literacy.



The degree of pre-existing knowledge about the health topic and the desire to obtain more information is also a user-oriented factor. Finally, there is a technological user-oriented factor, which is a general assessment of the users' access, preference, and frequency of use. Consistent with the TMC [3], the interactions between task- and user-oriented factors produce external stimuli, or "noise," which can serve as a hindrance or facilitator to the transaction.

Noise-inducing factors in the TMC, as well as other interactional communication models, comprise physical, psychological, physiological, and semantic factors [3,6,10]. Although evidence shows that diverse internal and external factors hinder and facilitate the capacity of eHealth users to successfully achieve their intended goals on the internet [48-50], no other eHealth literacy model or measurement instrument reviewed in this study considered the concept of "noise" as being part of the eHealth experience, beyond contextual eHealth factors. *Physical Noise* can include external factors that hinder the eHealth experience, including cognitive or information overload due to a wide variety of multimedia or physical challenges with the technology used (eg, screen size is too small or not bright enough, keyboards are too small). *Psychological Noise* includes the affective response to the eHealth experience, including the urgency for the information or the nature of the search (eg, cancer clinical trials vs physical activity information). *Physiological Noise* can be either temporal or permanent, meaning that it could be dexterity limitations due to a health condition or pain from a briefly debilitating migraine. Finally, *Semantic Noise* is the disagreement between meaning systems, including excessive use of scientific or wordy jargon from one or more communicators, as well as use of emojis or emoticons to transmit information. Ultimately, the degree of noise in a computer-mediated transaction is produced by the interaction between these task- and user-oriented factors.

The intrapersonal eHealth literacy (shown in Figure 2 as the hierarchical triangle) mediates the relationship between the effect of noise on eHealth contextual factors and the degree to which an eHealth end user is informed and empowered. Theoretically, eHealth literacy has an inverse relationship with the negative effect of noise. In other words, greater eHealth literacy negates the detrimental effects of noise produced from eHealth contextual factors and promotes a positive eHealth experience. This is consistent with the evidence that greater frequency of using eHealth improves proficiency in Web-based health information seeking [24] as users become more familiar with their information needs, the technology, and the usefulness of eHealth.

The intrapersonal skillset of eHealth literacy may be grounded in 3 foundational elements: Functional Literacy, Health Literacy [51], and Technology Literacy [52]. *Functional Literacy*, or the basic reading and writing skills [53], is a basic predecessor of both health and technology literacy. Together, health and technology literacy are central to eHealth skills [23]. The most recent definition of *Health Literacy* posits that it is [22]…

> *…linked to literacy and entails people's knowledge, motivation, and competence to access, understand, appraise, and apply health information in order to make judgments and take decisions in everyday life concerning health care, disease prevention, and health promotion to maintain or improve the quality of life during the life course.* [pg 3]

Technology Literacy is more concretely defined as "the ability to use, manage, assess, and understand technology" (pg 9) [52]. Without proficiency in functional literacy, eHealth users would not be able to successfully function within a health care context, let alone use technological devices or computer-mediated environments to address a health inquiry.

In our model, health and technology literacies shape a multidimensional and hierarchical intrapersonal skillset, which comprises 4 eHealth literacies. These eHealth literacies are aligned with the gold-standard health literacy model [53,54] to include *Functional*, *Communicative*, *Critical*, and *Translational eHealth Literacies*. Existing eHealth literacy definitions, models, and measurement instruments include a high volume of literacies and minimal insight into their relationship with underlying skillsets. Empirical evidence has hinted that the scientific community should consider reeling eHealth literacy conceptualizations back to seminal 4-tiered operational behaviors or literacies prevalent in health literacy and general literacy literature. This has been shown in a recent library of research on eHEALS, the seminal model of eHealth literacy, positing that it is not a unidimensional measure; rather, it may be a measure of 2-3 constructs that assess eHealth awareness, Web-based health information seeking skills, and evaluation or application skills [55-57]. The TMeHL builds on these principles to define 4 eHealth literacies.

The eHealth literacies presented in Figure 2 capture the hierarchical nature of these unique yet related skills, which map to operational skills proposed in our refined definition of eHealth literacy. Consistent with the health literacy literature [53,54], functional eHealth literacy is a foundational skill that precedes the remaining literacies. This literacy comprises lower-level operational skills, including the ability to locate and understand health information. Translational eHealth literacy is located at the highest level, as being proficient in this top-tier literacy requires a degree of proficiency to be present across all lower-level literacies. This hierarchical depiction shows that lower-level literacies and operational skills represent the necessary building blocks to achieve optimal proficiency in the higher-level literacies and operational eHealth skills. Stated differently, an eHealth lay end user must have basic skills in reading and writing and in typing to successfully exchange, evaluate, and apply health information from the internet. Each of these literacies is described and operationalized below, alongside the corresponding behaviors outlined in the proposed definition.

Functional eHealth Literacy (Operational Behaviors: to Locate and Understand)

According to Nutbeam [53], the definition of functional health literacy, which was adapted by Freebody and Luke [58], describes having:

> *Sufficient basic skills in reading and writing to be able to function effectively in everyday situations.* [pg 263]

Considering the technological context of functional health literacy, it is important to determine how well an individual can successfully read and write about health via a technological device. Therefore, *Functional eHealth Literacy* is defined as:

> *Basic skills in reading and writing (typing) about health to effectively function on the internet.*

Communicative eHealth Literacy (Operational Behavior: to Exchange)

Communicative literacy is [53]:

> *Advanced cognitive and literacy skills, which together with social skills, can be used to participate in everyday activities to extract information and derive meaning from different forms of communication and to apply new information to changing circumstances.* [pg 264]

In its original conceptualization, communicative literacy was intended to assess patients' communication skills when engaging with offline health care professionals [53,54]. eHealth is a computer-mediated form of communication, which provides limited salience to social and nonverbal cues [6]. According to Spitzberg and Cupach [59], success in achieving instrumental, self-presentation, and/or relational goals is determined based on the degree that interpersonal communication is appropriate and effective. Appropriate communication is consistent with social norms and relationships (stranger or close friend) among communicators. Effective communication helps achieve the desired goal of the health information seeking experience and interaction. There are 3 fundamental interpersonal communication skills that guide the degree to which someone is communicating appropriately and effectively [59]: (1) *control*, effectiveness in managing a situation to negotiate interpersonal problems and achieve a communicative goal; (2) *collaboration*, adhering to social norms to achieve an interaction goal; and (3) *adaptability*, acclimating to challenges by improving communicative styles based on contextual and social cues. These principles of interpersonal communication are also consistent with the underlying elements of participatory media that foster collaboration, openness, participation, and apomediation (Web 2.0), which differs from more static and linear, one-way information seeking behaviors (Web 1.0) [60]. As such, integrating interpersonal communication competence into eHealth literacy represents a unique contribution to understand the social aspects of eHealth. In our proposed model, *Communicative eHealth Literacy* is defined as:

> *The ability to collaborate, adapt, and control communication about health with users on social online environments with multimedia.*

Critical eHealth Literacy (Operational Behavior: to Evaluate)

Critical literacy is defined as [53]:

> *Advanced cognitive skills, which together with social skills, can be applied to critically analyze information and to use this information to exert greater control over life events and situations.* [pg 264]

Through the lens of the TMC, critical eHealth literacy includes being aware of the type of health information that is communicated to and from Web-based users, as well as the source from which this information is presented. This includes not only source and information credibility but also entails evaluating the relevance of and risks related to sharing personal information with Web-based sources through diverse channels. In this model, *Critical eHealth Literacy* is defined as:

> *The ability to evaluate the credibility, relevance, and risks of sharing and receiving health information on the internet.*

Translational eHealth Literacy (Operational Behavior: to Apply)

Developing a concept that acknowledges the dichotomy between "what people know" and "what people do" represents a fundamental gap in the health literacy literature [61]. Translating knowledge gained through a health-related interaction is the "process of moving what we learned…to the actual application of knowledge in a variety of practice settings and circumstances" [62]. In public health research, knowledge translation is a systems-level approach to transforming knowledge gained from rigorous research on societies for improved health outcomes

[62,63]. Within the context of eHealth, health information seekers often adopt the role of lay health researchers as they become exposed to new information including health-related knowledge from diverse sources (eg, peers, family, providers) and communication channels (eg, social media, electronic health records, news outlets). Strategies used to determine the applicability of new health information for translation into our existing knowledge structures depend on contextual factors, including personal and situational contexts [64]. This process likely depends on the skills a person has to identify and on the implementation of successful strategies for translating health information gained from electronic sources. Based on existing literature on knowledge translation, we propose the dimension "translational eHealth literacy." *Translational eHealth Literacy* is defined as:

> *The ability to apply health knowledge gained from the internet across diverse ecological contexts. Translational literacy is the highest cognitive level of eHealth literacy, meaning it is informed and built upon from all lower-level eHealth literacy dimensions (ie, critical, communicative, and functional).*

Discussion

Principal Findings

In this study, we systematically reviewed literature on eHealth literacy to provide an updated understanding of what we know about the construct in today's more transactional era of eHealth. Unlike Griebel et al [19], who suggested that eHealth literacy literature functions in solidarity and does not build upon prior literature, the results of this systematic review suggest that eHealth literacy literature has gradually built upon existing definitions and models to extend the construct to account for the evolving nature of eHealth. Unfortunately, while progress has been made, results from this concept analysis illustrate that existing literature's attempts continue to miss capturing the transactional nature of eHealth, specifically the skills needed to thrive within Web-based environments where social and contextual cues to action are limited. Instead, literature over the past decade has explored basic technological and contextual factors that influence individual eHealth literacy, but this work has provided little insight into transactional implications of eHealth literacy and intrapersonal skills that are important for cultivating positive eHealth experiences in various contexts. The intrapersonal skillset of eHealth literacy remains underdeveloped, especially regarding the role of communication. Results from this study were used to present a theoretical proposition of eHealth literacy that supports the transactional elements of eHealth. Subsequently, this new knowledge is leveraged to generate a refined definition of eHealth literacy and its complementary model.

Existing eHealth literacy definitions include operational skills required for an eHealth end user to thrive on the internet (ie, locate, understand, evaluate, apply, and, most recently, communicate or create). However, dimensions of existing models and measures are not intuitively aligned with the intrapersonal operational skills outlined in their corresponding definitions. Norman and Skinner's [15] eHEALS was intended to serve as a unidimensional scale to capture the Lily Model's eHealth literacy, or the self-efficacy to locate, understand, evaluate, and act upon Web-based health resources [14]. Over the past decade, strong empirical evidence has shown that eHEALS is a three-dimensional measure that assesses eHealth users' self-efficacy in their eHealth awareness, information seeking, and evaluation and actions related to Web-based health information [55-57]. This research begins to clarify the relationship between operational skills outlined in eHealth literacy definitions and dimensions captured in corresponding measurement instruments. Moreover, limited empirical attention has been paid to the transactional operational skills needed to thrive within the social era of eHealth. The proposed TMeHL seeks to bridge this fundamental disconnect in the eHealth literacy literature by proposing a definition and model that specify important operational skills and literacies that should be considered.

The dimension of "communication" was significantly underdeveloped in eHealth literacy definitions, models, and measurement instruments reviewed. "Communication" was not integrated within definitions until 2011, and it first appeared in a conceptual model in 2014 as an outcome related to high eHealth literacy, not as an integral or defining element [26]. Communication was not considered a core element of eHealth literacy; rather, existing measures stressed the importance of "interactivity," or the ability to talk about findings from an internet search with an offline health care provider [32,34]. The most recent measurement instrument operationalizes "communication" as the ability to self-create, add, or generate messages on social media with a technological device [35]. Interestingly, in the most recent definition, Griebel et al [19] posited that "communicating" and "creating" are two discrete skills. The role of communication appears to be having an identity crisis in eHealth literacy literature. In the TMC, particularly in computer-mediated contexts, communication is a vehicle that facilitates the process of cocreating information within diverse contexts among two or more communicators [3,5]. The proposed TMeHL definition and model consider communication as a central skill of eHealth literacy that affects critical (evaluative) and translational (application) elements of the eHealth experience. Researchers should view communication in this manner, rather than as an end goal or single act of generating a health-related message.

Limitations

This study is not without limitations. Despite the rigorous extraction procedures and inclusion of gray literature, it is possible that not all eHealth literacy models, definitions, and measures were included because of the time frame of our literature extraction procedures. A qualitative approach was used to extract and analyze the literature in this study, which is prone to researcher bias [65]. However, we applied a concept analysis method [20,21], which is a rigorous and well-regarded approach to refine and operationalize an evolving concept, like eHealth literacy. This study proposes a refined definition and model of eHealth literacy to assist researchers and practitioners in "keeping up" with the evolving nature and transactional approaches currently ingrained in eHealth. We present a theoretical proposition of eHealth literacy that supports the transactional elements of eHealth.

The theoretical tenants from TMC that were used to derive the TMeHL were informed by results of previously published systematic reviews, theory-driven articles, measurement development studies, and empirically driven original research articles examining eHealth literacy in diverse populations. Published research informed development of TMeHL, but the model has not been subjected to any formal evaluation or hypothesis testing. Per recommendations by Griebel et al [19], future research is needed to obtain key stakeholder feedback about eHealth literacy models. Formative validation based on stakeholder input will provide an additional layer of validity evidence to support model testing in quantitative studies. Specifically, there is a need to explore how eHealth literacy serves as a mediator to counteract the negative effects of noise in eHealth transactions, as described in this study.

Conclusion

Existing eHealth literacy definitions, models, and measures do not account for the transactional nature of eHealth. Few have sufficient theoretical underpinnings. Prior to our contribution, researchers had yet to capture the "social" elements of eHealth with theoretical underpinnings from the perspective of transactional communication. This is primarily due to the high volume of overlapping and inconsistent literacies, as well as the underdeveloped nature of "communication" as an integral component of eHealth literacy. In addition, existing eHealth literacy definitions, models, and measures failed to adequately integrate "communication" as an essential component of eHealth literacy. Because of this, a refined eHealth literacy definition and model based on the TMC are proposed. There is a need to validate this model with key stakeholders in eHealth and test the assumptions of the model with eHealth experts and lay end users.

Acknowledgments

Research reported in this publication was supported by the National Heart, Lung and Blood Institute of the National Institutes of Health under award #F31HL132463. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Dimensions (content areas) of eHealth literacy conceptual models and measures.

[PDF File (Adobe PDF File), 44KB - jmir_v20i10e10175_app1.pdf]

References

1. Barello S, Triberti S, Graffigna G, Libreri C, Serino S, Hibbard J, et al. eHealth for Patient Engagement: A Systematic Review. Front Psychol 2015;6:2013 [FREE Full text] [doi: 10.3389/fpsyg.2015.02013] [Medline: 26779108]
2. Shaw T, McGregor D, Brunner M, Keep M, Janssen A, Barnet S. What is eHealth (6)? Development of a Conceptual Model for eHealth: Qualitative Study with Key Informants. J Med Internet Res 2017 Oct 24;19(10):e324 [FREE Full text] [doi: 10.2196/jmir.8106] [Medline: 29066429]
3. Barnlund D. Foundations of Communication Theory. Communication: The context of change. In Sereno KK, Mortensen CD. Foundations of communication theory, pp. 83-102. Harper & Row; New York, USA; 1970.
4. Walther JB, Jang J. Communication Processes in Participatory Websites. J Comput-Mediat Comm 2012 Oct 10;18(1):2-15. [doi: 10.1111/j.1083-6101.2012.01592.x]
5. Knapp M, Daly J. The SAGE Handbook of Interpersonal Communication. In: 4th ed. SAGE Publications, Inc. Thousand Oaks, CA: SAGE Publications; 2011.
6. DeVito J. The Interpersonal Communication Book. 14th edition. Boston MA: Pearson; 2013:0133753816.
7. Ramirez A, Walther J, Burgoon J, Sunnafrank M. Information-seeking strategies, uncertainty, and computer-mediated communication. Hum Commun Res 2002;28(2):213-228.
8. Hargittai E. Digital na(t)ives? Variation in internet skills and uses among members of the ?Net Generation. ? Sociol Inq ? 2010;80:113. [doi: 10.1111/j.1475-682X.2009.00317.x]
9. Lupiáñez-Villanueva F, Mayer MA, Torrent J. Opportunities and challenges of Web 2.0 within the health care systems: an empirical exploration. Inform Health Soc Care 2009 Sep;34(3):117-126. [doi: 10.1080/17538150903102265] [Medline: 19670002]
10. Hargie O. Skilled Interpersonal Communication: Research, Theory and Practice. 5th edition. New York, USA: Routledge; 2011:0415432049.
11. van Deursen AJAM, van Dijk JAGM. Internet skills performance tests: are people ready for eHealth? J Med Internet Res 2011 Apr;13(2):e35 [FREE Full text] [doi: 10.2196/jmir.1581] [Medline: 21531690]

12. Silver MP. Patient perspectives on online health information and communication with doctors: a qualitative study of patients 50 years old and over. J Med Internet Res 2015;17(1):e19 [FREE Full text] [doi: 10.2196/jmir.3588] [Medline: 25586865]
13. McCloud RF, Okechukwu CA, Sorensen G, Viswanath K. Beyond access: barriers to internet health information seeking among the urban poor. J Am Med Inform Assoc 2016 Nov;23(6):1053-1059 [FREE Full text] [doi: 10.1093/jamia/ocv204] [Medline: 27206459]
14. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. J Med Internet Res 2006 Jun;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
15. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
16. Bautista J. From solving a health problem to achieving quality of life: redefining eHealth literacy. J Lit Technol 2015;16(2):54. [Medline: 286688300]
17. Norman C. eHealth literacy 2.0: problems and opportunities with an evolving concept. J Med Internet Res 2011 Dec;13(4):e125 [FREE Full text] [doi: 10.2196/jmir.2035] [Medline: 22193243]
18. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
19. Griebel L, Enwald H, Gilstad H, Pohl A, Moreland J, Sedlmayr M. eHealth literacy research-Quo vadis? Inform Health Soc Care 2017 Oct 18:1-16. [doi: 10.1080/17538157.2017.1364247] [Medline: 29045164]
20. Nuopponen A. Methods of concept analysis - a comparative study. Lang Spec Purp Prof Commun Knowl Manag Cogn 2010;1(1):4-15 [FREE Full text]
21. Walker L, Avant K. Strategies for Theory Construction in Nursing. 5th edition. Boston: Pearson; 2010.
22. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z. Health literacy and public health: A systematic review and integration of definitions and models. BMC Public Health 2012;12:A. [doi: 10.1186/1471-2458-12-80]
23. Bodie GD, Dutta MJ. Understanding health literacy for strategic health marketing: eHealth literacy, health disparities, and the digital divide. Health Mark Q 2008 Jan;25(1-2):175-203. [doi: 10.1080/07359680802126301] [Medline: 18935884]
24. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
25. Chan CV, Kaufman DR. A framework for characterizing eHealth literacy demands and barriers. J Med Internet Res 2011 Nov;13(4):e94 [FREE Full text] [doi: 10.2196/jmir.1750] [Medline: 22094891]
26. Gilstad H. CEUR Workshop Proceedings. Toward a comprehensive model of eHealth literacy URL: http://ceur-ws.org/Vol-1251/paper7.pdf [accessed 2018-02-18] [WebCite Cache ID 6xKLH9H1V]
27. Paek H, Hove T. Social cognitive factors and perceived social influences that improve adolescent eHealth literacy. Health Commun 2012;27(8):727-737. [doi: 10.1080/10410236.2011.616627] [Medline: 22452551]
. Brynn R. Universal Design and Welfare Technology. Stud Health Technol Inform 2016;229:335-344. [Medline: 27534324]
29. Werts N, Hutton-Rogers L. Barriers To Achieving E-Health Literacy. AJHS 2013 Aug 14;4(3):115. [doi: 10.19030/ajhs.v4i3.8007]
30. Kayser L, Kushniruk A, Osborne RH, Norgaard O, Turner P. Enhancing the Effectiveness of Consumer-Focused Health Information Technology Systems Through eHealth Literacy: A Framework for Understanding Users' Needs. JMIR Hum Factors 2015 May 20;2(1):e9 [FREE Full text] [doi: 10.2196/humanfactors.3696] [Medline: 27025228]
31. Chew F, Yuqian N. Promoting health among older adults via eHealth literacy. 2015 Presented at: Proceedings of the International Conference on Communication, Media, Technology, and Design; May 2015; Dubai, United Arab Emirates p. 16-18.
32. Koopman RJ, Petroski GF, Canfield SM, Stuppy JA, Mehr DR. Development of the PRE-HIT instrument: patient readiness to engage in health information technology. BMC Fam Pract 2014;15:18 [FREE Full text] [doi: 10.1186/1471-2296-15-18] [Medline: 24472182]
33. Bhalla A, Durham RL, Al-Tabaa N, Yeager C. The development and initial psychometric validation of the eHealth readiness scale. Computers in Human Behavior 2016 Dec;65:460-467. [doi: 10.1016/j.chb.2016.09.015]
34. Seçkin G, Yeatts D, Hughes S, Hudson C, Bell V. Being an Informed Consumer of Health Information and Assessment of Electronic Health Literacy in a National Sample of Internet Users: Validity and Reliability of the e-HLS Instrument. J Med Internet Res 2016 Jul 11;18(7):e161 [FREE Full text] [doi: 10.2196/jmir.5496] [Medline: 27400726]
35. van der Vaart R, Drossaert C. Development of the Digital Health Literacy Instrument: Measuring a Broad Spectrum of Health 1.0 and Health 2.0 Skills. J Med Internet Res 2017 Jan 24;19(1):e27 [FREE Full text] [doi: 10.2196/jmir.6709] [Medline: 28119275]
36. Milne RA, Puts MTE, Papadakos J, Le LW, Milne VC, Hope AJ, et al. Predictors of High eHealth Literacy in Primary Lung Cancer Survivors. J Cancer Educ 2015 Dec;30(4):685-692. [doi: 10.1007/s13187-014-0744-5] [Medline: 25355524]
37. Chang F, Chiu C, Chen P, Miao N, Lee C, Chiang J, et al. Relationship Between Parental and Adolescent eHealth Literacy and Online Health Information Seeking in Taiwan. Cyberpsychol Behav Soc Netw 2015 Oct;18(10):618-624. [doi: 10.1089/cyber.2015.0110] [Medline: 26375050]

38. Chen W, Lee K. More than search? Informational and participatory eHealth behaviors. Computers in Human Behavior 2014 Jan;30:103-109. [doi: 10.1016/j.chb.2013.07.028]
39. Paige SR, Krieger JL, Stellefson ML. The Influence of eHealth Literacy on Perceived Trust in Online Health Communication Channels and Sources. J Health Commun 2017 Jan;22(1):53-65. [doi: 10.1080/10810730.2016.1250846] [Medline: 28001489]
40. Mitsutake S, Shibata A, Ishii K, Oka K. Association of eHealth literacy with colorectal cancer knowledge and screening practice among internet users in Japan. J Med Internet Res 2012 Nov 13;14(6):e153 [FREE Full text] [doi: 10.2196/jmir.1927] [Medline: 23149453]
41. Choi NG, Dinitto DM. The digital divide among low-income homebound older adults: Internet use patterns, eHealth literacy, and attitudes toward computer/Internet use. J Med Internet Res 2013 May;15(5):e93 [FREE Full text] [doi: 10.2196/jmir.2645] [Medline: 23639979]
42. Sheng X, Simpson PM. Seniors, health information, and the Internet: motivation, ability, and Internet knowledge. Cyberpsychol Behav Soc Netw 2013 Oct;16(10):740-746. [doi: 10.1089/cyber.2012.0642] [Medline: 23679569]
43. Watkins I, Xie B. eHealth literacy interventions for older adults: a systematic review of the literature. J Med Internet Res 2014;16(11):e225 [FREE Full text] [doi: 10.2196/jmir.3318] [Medline: 25386719]
44. 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]
45. Hsu W, Chiang C, Yang S. The effect of individual factors on health behaviors among college students: the mediating effects of eHealth literacy. J Med Internet Res 2014;16(12):e287 [FREE Full text] [doi: 10.2196/jmir.3542] [Medline: 25499086]
46. Kayser L, Karnoe A, Furstrand D, Batterham R, Christensen KB, Elsworth G, et al. A Multidimensional Tool Based on the eHealth Literacy Framework: Development and Initial Validity Testing of the eHealth Literacy Questionnaire (eHLQ). J Med Internet Res 2018 Feb 12;20(2):e36 [FREE Full text] [doi: 10.2196/jmir.8371] [Medline: 29434011]
47. Nørgaard O, Furstrand D, Klokker L, Karnoe A, Batterham R, Kayser L. The e-health literacy framework: A conceptual framework for characterizing e-health users and their interaction with e-health systems. Knowl Manag E-Learn Int J 2015;7:40.
48. Kim E, Stolyar A, Lober WB, Herbaugh AL, Shinstrom SE, Zierler BK, et al. Challenges to using an electronic personal health record by a low-income elderly population. J Med Internet Res 2009 Oct;11(4):e44 [FREE Full text] [doi: 10.2196/jmir.1256] [Medline: 19861298]
49. Mirkovic J, Kaufman DR, Ruland CM. Supporting cancer patients in illness management: usability evaluation of a mobile app. JMIR Mhealth Uhealth 2014;2(3):e33 [FREE Full text] [doi: 10.2196/mhealth.3359] [Medline: 25119490]
50. van der Vaart R, Drossaert CHC, de Heus M, Taal E, van de Laar MAFJ. Measuring actual eHealth literacy among patients with rheumatic diseases: a qualitative analysis of problems encountered using Health 1.0 and Health 2.0 applications. J Med Internet Res 2013 Feb;15(2):e27 [FREE Full text] [doi: 10.2196/jmir.2428] [Medline: 23399720]
51. National Center for Education Statistics. National Assessment of Adult Literacy (NAAL): Three types of literacy URL: https://nces.ed.gov/naal/literacytypes.asp [accessed 2018-02-18] [WebCite Cache ID 6xKcQYo9e]
52. International Technology Education and Association: Content for the Study of Technology. 2007 URL: https://www.iteea.org/File.aspx?id=67767&v=b26b7852 [WebCite Cache ID 6xKcaq1vB]
53. Nutbeam D. Health literacy as a public health goal: A challenge for contemporary health education and communication strategies into the 21st century. Health Promot Int 2000;15(3):259-267. [doi: 10.1093/heapro/15.3.259]
54. Nutbeam D. The evolving concept of health literacy. Soc Sci Med 2008 Dec;67(12):2072-2078. [doi: 10.1016/j.socscimed.2008.09.050] [Medline: 18952344]
55. Soellner R, Huber S, Reder M. The Concept of eHealth Literacy and Its Measurement. Journal of Media Psychology 2014 Jan;26(1):29-38. [doi: 10.1027/1864-1105/a000104]
56. Sudbury-Riley L, FitzPatrick M, Schulz PJ. Exploring the Measurement Properties of the eHealth Literacy Scale (eHEALS) Among Baby Boomers: A Multinational Test of Measurement Invariance. J Med Internet Res 2017 Feb 27;19(2):e53 [FREE Full text] [doi: 10.2196/jmir.5998] [Medline: 28242590]
57. Neter E, Brainin E, Baron-Epel O. The dimensionality of health literacy and eHealth literacy. Eur Health Psychol 2015;17:80 [FREE Full text]
58. Freebody L, Luke A. Literacies programs: Debates and demands in cultural context. Prospect Aust Jouranl TESOL 1990;5(3):7-16 [FREE Full text]
59. Spitzberg B, Cupach W. Interpersonal Communication Competence. University of Michigan: SAGE Publications; 1984.
60. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008;10(3):e22 [FREE Full text] [doi: 10.2196/jmir.1030] [Medline: 18725354]
61. National Academy of Medicine. Considerations for a New Definition of Health Literacy URL: https://nam.edu/considerations-for-a-new-definition-of-health-literacy/ [accessed 2018-02-18] [WebCite Cache ID 6xKcjYBXP]
62. Center on Knowledge Translation for Disability and Rehabilitation Research. Knowledge Translation: Introduction to Models, Strategies, and Measures URL: http://ktdrr.org/ktlibrary/articles_pubs/ktmodels/index.html [accessed 2018-02-18] [WebCite Cache ID 6xKcxJykY]

63. Woolf SH. The meaning of translational research and why it matters. JAMA 2008 Jan 9;299(2):211-213. [doi: 10.1001/jama.2007.26] [Medline: 18182604]
64. Straus SE, Tetroe J, Graham I. Defining knowledge translation. Canadian Medical Association Journal 2009 Jul 20;181(3-4):165-168. [doi: 10.1503/cmaj.081229]
65. Anderson C. Presenting and evaluating qualitative research. Am J Pharm Educ 2010 Oct 11;74(8):141 [FREE Full text] [Medline: 21179252]

Abbreviations

eHealth: electronic health
eHEALS: Electronic Health Literacy Scale
I-CMC: interpersonal computer-mediated communication
TMC: Transactional Model of Communication
TMeHL: Transactional Model of eHealth Literacy

Edited by G Eysenbach; submitted 20.02.18; peer-reviewed by E Brainin, E Neter, K Spanjers, L Kayser; comments to author 20.04.18; revised version received 18.05.18; accepted 28.06.18; published 02.10.18.

Please cite as:
Paige SR, Stellefson M, Krieger JL, Anderson-Lewis C, Cheong J, Stopka C
Proposing a Transactional Model of eHealth Literacy: Concept Analysis
J Med Internet Res 2018;20(10):e10175
URL: https://www.jmir.org/2018/10/e10175/
doi:10.2196/10175
PMID:30279155

Original Paper

Associations of eHealth Literacy With Health Services Utilization Among College Students: Cross-Sectional Study

Yi Fang Luo[1*]; Shu Ching Yang[1*], PhD; An-Sing Chen[2*], PhD; Chia-Hsun Chiang[1*], PhD

[1]Institute of Education, National Sun Yat-sen University, Kaohsiung, Taiwan

[2]Department of Finance, National Chung Cheng University, Chiayi, Taiwan

[*]all authors contributed equally

Corresponding Author:
Chia-Hsun Chiang, PhD
Institute of Education
National Sun Yat-sen University
70 Lien-hai Road
Kaohsiung, 80424
Taiwan
Phone: 886 47612346
Email: d996050002@student.nsysu.edu.tw

Abstract

Background: Electronic health (eHealth) literacy has become an important topic in health fields. Studies have found that individuals with higher eHealth literacy are more likely to use preventive care services and to have effective interactions with their physicians. In addition, previous studies have revealed a gender difference in the utilization of physician access and outpatient services. Nevertheless, few studies have explored the effect of the three levels of eHealth literacy (functional, interactive, and critical levels) on the four aspects of health services utilization (type, site, purpose, and time interval). It is unclear whether the associations between these three levels of eHealth literacy and the four aspects of health services utilization among college students are positive or negative.

Objective: The objective of this study was to investigate the associations among gender, eHealth literacy, and health services utilization.

Methods: We used the eHealth Literacy Scale, a 12-item instrument designed to measure college students' functional, interactive, and critical eHealth literacy, and the Health Services Utilization Scale, which is a 10-item instrument developed to measure the four aspects of health services utilization by college students. A nationally representative sample of 489 college students in Taiwan was surveyed. We conducted multiple regression analysis to examine the associations among gender, eHealth literacy, and health services utilization.

Results: The study found that being female was negatively related to the purpose aspect of health services utilization ($t_{487}=-2.85$, $P<.01$). However, the R^2 value of gender on the purpose aspect was low enough to be ignored. Critical ($t_{484}=2.98\text{-}4.23$, $P<.01$) and interactive eHealth literacy ($t_{484}=2.43\text{-}2.89$, $P<.05$) were related to three aspects of the health services utilization, and functional eHealth literacy was related to the purpose aspect ($t_{484}=-4.99$, $P<.001$).

Conclusions: This study showed that Taiwanese college students with interactive eHealth literacy were more likely to have a higher rate of outpatient care use. Moreover, Taiwanese college students with critical eHealth literacy were more likely to make full use of health services than those with functional eHealth literacy. Finally, the educated and age-restricted sample may attenuate gender disparities in health services utilization among Taiwanese college students.

(J Med Internet Res 2018;20(10):e283) doi:10.2196/jmir.8897

KEYWORDS

eHealth; health literacy; health services; student; utilization

Introduction

Health Literacy and eHealth Literacy

In the past, individuals were likely to obtain health information from written sources such as books, magazines, newspapers, or brochures. The advent of the internet has drastically changed health information, and the internet is widely used to obtain this information. However, the skills needed to collect information through the internet differ from those needed to collect information from books and leaflets. People need to not only be health literate but also have capabilities, resources, and motivation to find, understand, and appraise health information when using digital services and technology [1]. Therefore, to obtain a complete overview of people's skills in obtaining and using health information, it is more necessary to measure electronic health (eHealth) literacy than to measure health literacy [2].

eHealth literacy is an extension of the health literacy concept. Health literacy refers to an individual's ability to understand health care information and to make appropriate decisions [3]. eHealth literacy, which consists of skills related to health literacy and digital literacy [1], is defined as the ability to seek, find, understand, and appraise health information from electronic sources and to apply the knowledge gained to address or solve health problems [4]. Based on Nutbeam's concept [5], eHealth literacy includes functional, interactive, and critical levels. At the most basic level, functional eHealth literacy refers to basic reading and writing skills and basic knowledge of health conditions and health systems. Interactive eHealth literacy refers to communicative and social skills that can be used to abstract information and derive meaning from different forms of communication. The highest level of critical eHealth literacy builds on functional and interactive literacy and involves the most advanced cognitive skills that can be applied to critically analyze information, discern the quality of health websites, and use good information to make informed decisions about health [6,7].

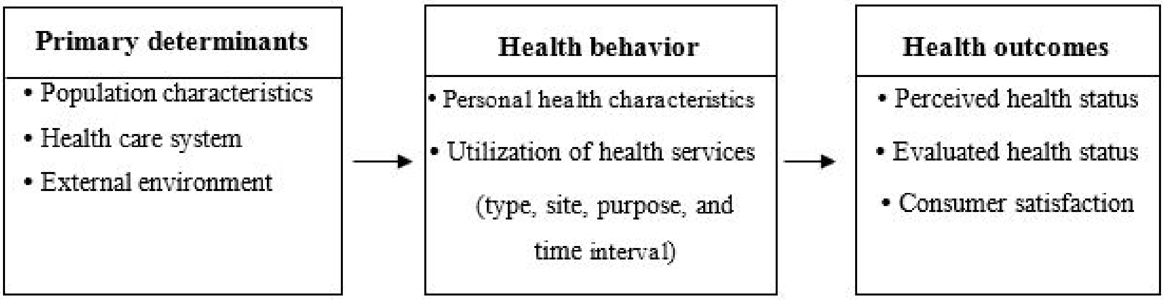
eHealth literacy has become an important topic in the health fields. Studies have found that individuals with high eHealth literacy are more likely to adopt healthy behaviors [6,8] and demonstrate better health responsibility and self-actualization [7]. Recently, access to health services and the relational dynamics between patients and health care providers have been transformed by Web-based health information-seeking behaviors [9] and eHealth literacy [10]. Therefore, the aim of this study is to examine the extent to which functional, interactive, and critical eHealth literacy are associated with college students' utilization of health services.

Health Services Utilization

Andersen [11] suggests four relevant aspects of the utilization of health services, each reflecting different aspects of the care-seeking process (Figure 1). The type of utilization indicates what kind of service is obtained and who provides it (eg, hospital, physician, dentist, or pharmacist). The site of the medical care encounter is the place where the care is received (eg, physician's office, hospital clinic, or emergency room). The purpose of the visit signifies the reasons for care-seeking such as a need for preventive, illness-related, or custodial care. The time interval for a visit can be represented by contact, volume, or continuity measures [12].

Gender has been identified as a predisposing factor of health services utilization [13-16]. As increasing numbers of studies investigate the relationship between gender and health services utilization in different contexts, important inconsistencies have been reported. For example, some studies have found that females are more likely than males to visit physicians [17] and tend to use more outpatient services [18]. Studies have also found that females perceive a greater need for mental health treatment [19] and have a higher tendency to use all services, except informal providers, than males [20]. However, other studies have shown that females suffering from illness report seeking health care less frequently than men [21]. Females were found to be more likely to report discrimination in health care and less likely to receive preventive health services [22]. Furthermore, other studies have found no statistically significant evidence of a relationship [23,24]. College students are an educated and transition-age population group. They are in the upper part of the distribution in terms of their resources. Previous studies have indicated that the combination of education [25] and women's autonomy [26] appear to attenuate gender disparities in health services utilization. Therefore, we want to investigate whether there are gender differences in the four aspects of health services utilization among college students.

Figure 1. Andersen model of health services utilization.



Relationship Between eHealth Literacy and the Utilization of Health Services

Individuals with limited health literacy are at an increased risk of poorer health outcomes. Systematic reviews have demonstrated that low health literacy is consistently associated with more hospitalizations, greater use of emergency care, and lower receipt of mammography screening and influenza vaccine [27]. In contrast, higher health literacy supports effective involvement in care processes. Studies have shown that individuals with adequate health literacy are more likely to be satisfied with care [28] and use preventive health services [29].

Web-based health information-seeking behaviors and eHealth literacy are transforming the physician-patient relationship and health services [9,10]. People who use eHealth resources (eg, health websites) are more likely to be well prepared for clinical visits, ask more related questions, learn more about their health care, and take actions to enhance their health than those who do not [30]. Studies have shown that internet use is associated with increased use of health care services and better-perceived outcomes of internet use for health purposes [31], including the decision to use health services and the way individuals communicate with physicians and request referrals to specialist care [32]. In addition, studies have shown that individuals with higher eHealth literacy are more likely to have knowledge of and a screening history for colorectal cancer [33] and more effective interactions with their physicians [34]. Moreover, some studies have found a set of indirect effect paths, via health information seeking and empowerment, showing a positive relationship between eHealth literacy and health care utilization [35]. However, these studies did not examine the effect of the three levels of eHealth literacy on the four aspects of medical services usage. Therefore, the research question examined for this study is as follows: Are the associations between the three levels of eHealth literacy and the four aspects of health services utilization among college students positive or negative?

Methods

Participants

Pretest Sample

Pretesting can help authors identify questions that do not make sense to participants as well as problems with questionnaires that might lead to biased answers. Thus, pretesting was conducted to develop and test the adequacy of the research instrument designed by the authors. Confirmatory factor analysis (CFA) was used to assess the reliability of the survey instrument. A reasonable sample size for a simple CFA model is approximately N=100-150 [36]. For this study, a purposive sample of 240 college students was drawn from 1 comprehensive university in Taiwan. Each participant was mailed a questionnaire, and 193 usable (completed) questionnaires were returned, resulting in an effective response rate of 80.4% (193/240).

Formal Study Sample

This cross-sectional study was conducted in Taiwan. We recruited 600 college students from 9 schools to participate in the survey. Teachers in selected colleges were contacted to request their assistance in the questionnaire distribution. Of the 529 recovered surveys, 40 invalid surveys (respondents had not completed the entire survey or gave invalid responses) and 489 valid surveys (92.4%, 489/529) were retained. The number of valid surveys was >350 [37]; thus, we deleted the invalid surveys. Among these 489 valid respondents (183 male and 306 female), 28.0% (137/489) studied in the northern region, 22.3% (109/489) studied in the central region, and 49.7% (243/489) studied in the southern region of Taiwan. The participants' mean (SD) age was 21.51 (4.11) years. Of the 489 participants, 85.3% (417/489) were aged <22 years.

Survey Instrument

eHealth Literacy Scale

eHealth Literacy was assessed by Chiang et al's eHealth Literacy Scale (eHLS) [8]. The validity and reliability of Chiang et al's eHLS were tested using item analyses, exploratory factor analysis, and CFA, which revealed that the eHLS is a reliable and validated measure of functional, interactive, and critical eHealth literacy for Taiwan college students. The 12-item eHLS assesses functional (3 items, eg, the Web-based health information is too difficult to understand), interactive (4 items, eg, paying attention to Web-based health information), and critical (5 items, eg, evaluating the effectiveness and reliability of Web-based health information) eHealth literacy. Responses were given on a 5-point Likert scale ranging from 5 (total agreement) to 1 (total disagreement). Mean scores for the eHLS were calculated by summing the item scores divided by the total number of items, resulting in a score ranging from 1 (lower eHealth literacy) to 5 (higher eHealth literacy).

Amos 6.0 CFA was used to examine the best measurement model. An analysis was conducted with Amos using maximum likelihood estimation. The factor loading of the 12-item eHLS ranged from 0.61 to 0.86. The individual item reliability of the 12-item eHLS ranged from 0.38 to 0.74.

The composite reliability ranged from 0.80 to 0.91, and the average variance extracted for each level ranged from 0.57 to 0.66. The cutoff values of CFA included a chi-square/degree of freedom value of <3, a root mean square residual (RMR) value of <0.50, a standardized root mean square residual value (SRMR) of <0.50, a root mean square error of approximation (RMSEA) value of <0.08, a goodness of fit index (GFI) value of >0.90, an adjusted GFI (AGFI) value of >0.90, a normed fit index (NFI) value of >0.90, a Tucker-Lewis index (TLI) value of >0.90, a relative fit index (RFI) value of >0.90, an incremental fit index (IFI) value of >0.90, a comparative fit index (CFI) value of >0.90, a parsimonious goodness of fit index (PGFI) value of >0.50, a parsimonious normed fit index (PNFI) value of >0.50, and a parsimonious comparative fit index (PCFI) value of >0.50 [38,39]. We found that the current data adequately fit the eHLS model, which was divided into 3 levels (total 12 items). We used Amos to conduct CFA, and a review of the fit indices revealed a chi-square/degree of freedom value of 1.68, an RMR value of 0.02, an SRMR value of 0.40, an RMSEA value of 0.06, a GFI value of 0.93, an AGFI value of 0.90, an NFI value of 0.94, a TLI value of 0.97, an RFI value of 0.92, an IFI value of 0.97, a CFI value of 0.97, a PGFI value of 0.61, a PNFI value of 0.72, and a PCFI value of 0.75. Furthermore, the chi-square test was significant (χ^2_{51}=85.50, *P*=.002). However, the chi-square statistic is, in essence, a statistical significance test that is sensitive to sample size, which means that this statistic nearly always rejects the model when large samples are used [38,40].

Health Services Utilization Scale

We developed the Health Services Utilization Scale (HSUS) following a thorough review of the literature [12,23] and applying the process of concept clarification. The content validity of the HSUS was pretested by three specialist professors. It was also pretested with 5 college students. It contains the following four dimensions (Multimedia Appendix 1):

1. Type: The tendency to make good use of various kinds of health care service organizations (3 items, eg, receiving various forms of medical treatment such as Chinese medicine, Western medicine, and dental services).
2. Site: The tendency to make good use of a multitiered health care system (2 items, eg, choosing a suitable site for medical services such as a physician's office, hospital clinic, or emergency room).
3. Purpose: The ability to seek medical advice based on different needs (3 items, eg, visiting a physician for preventive care).
4. Time interval: The frequency of medical use (2 items, eg, obtaining a second opinion from another physician).

The items were answered using a 5-point Likert scale with scores ranging from 1 (never) to 5 (always). High scores in the respective aspects indicated a greater likelihood of the respondents making good use of various kinds of health care services and a multitiered health care system, seeking medical advice based on different needs, and having a greater frequency and ratio of outpatient care use.

The current data had a good fit with the model, which was divided into 4 aspects (10 total items). We used Amos to conduct CFA and found that the factor loading of the 10-item HSUS ranged from 0.66 to 0.82. The individual item reliability of the 10-item HSUS ranged from 0.43 to 0.68. The composite reliability ranged from 0.72 to 0.77, and the average variance extracted for each aspect ranged from 0.51 to 0.58. A review of the fit indices revealed a chi-square/degree of freedom value of 2.12, an RMR value of 0.05, an SRMR value of 0.50, an RMSEA value of 0.08, a GFI value of 0.94, an AGFI value of 0.89, an NFI value of 0.91, a TLI value of 0.92, an RFI value of 0.90, an IFI value of 0.95, a CFI value of 0.95, a PGFI value of 0.50, a PNFI value of 0.59, and a PCFI value of 0.61. Furthermore, the chi-square test was significant (χ^2_{29}=61.42, *P*<.001). According to the cutoff values of CFA [38,39], the indicators demonstrated a good fit for the measurement model.

Data Analysis

First, peer review was used to confirm the content validity of the HSUS. Second, we used Amos 6.0 (IBM Corp, Armonk, NY, USA) to perform CFA to identify the best measurement models for the eHLS and HSUS. Finally, SPSS 18.0 (IBM Corp, Armonk, NY, USA) was used to conduct hierarchical multiple regression analysis. We performed four hierarchical multiple regression analyses to examine the explanatory power of the four aspects of health services utilization. We determined the order in which the variables were entered into the model based on logical or theoretical considerations. Thus, gender was entered in step 1, and the three levels of eHealth literacy were entered in step 2.

Ethical Considerations

The study was reviewed and approved by the Institute of Education at the National Sun Yat-Sen University, Taiwan. The study adopted an anonymous questionnaire, in line with the government's institutional review board rules for exempt review. The questionnaire instructions informed the participants of the research purpose and confidentiality and indicated that they had the right to refuse to participate at any time. The participants received the questionnaire and gifts at the same time; even if a participant decided to drop out of the investigation, he or she still received the gifts (a pen and an L-folder). This approach was intended to be fair to each participant, to avoid the impact of gift incentives on the participants, and to provide compensation for the participants.

Results

Descriptive Statistics of eHealth Literacy and Health Services Utilization

Among all participants, the mean score of functional eHealth literacy was 3.66 (SD 0.70), of interactive eHealth literacy was 3.67 (SD 0.67), and of critical eHealth literacy was 3.65 (SD 0.69), indicating that college students had medium or above levels of eHealth literacy.

In HSUS, the mean score of the type aspect was 3.36 (SD 0.75), of site was 3.32 (SD 0.77), of purpose was 3.70 (SD 0.81), and of time interval was 2.37 (SD 0.80). This result indicated that college students had a certain degree of ability to make good use of various kinds of health care services and a multitiered health care system and to receive health care services based on different needs, although they did not frequently use outpatient care.

Analysis of Gender and Health Services Utilization

In the analysis of Model 1 (see Multimedia Appendix 2), male and female groups were transformed into dummy variables. We used the male group as the reference group. Multimedia Appendix 2 indicates that membership in the female group was negatively related to the purpose aspect of health services utilization, yielding a low explanatory power of 2%. However, the R^2 value is very low to nil.

Analysis of Gender, eHealth Literacy, and Health Services Utilization

When comparing Models 1 and 2 (see Multimedia Appendix 3), the level of explanatory power for the 4 health services utilization aspects increases by 6%-16%. Moreover, Multimedia Appendix 3 shows that when controlling for gender, interactive and critical eHealth literacy are both positively related to three aspects of health services utilization. Functional eHealth literacy is negatively related to the time interval aspect.

Discussion

Principal Findings

This study found no gender difference in health services utilization among Taiwanese college students. There was a statistically significant correlation between eHealth literacy and health services utilization. Finally, among the three eHealth literacy levels, functional eHealth literacy showed the lowest correlation to health services utilization.

The study found no significant gender differences in health utilization. Notably, the R^2 value of gender on the purpose aspect was low enough to be ignored. The association between gender and service utilization has been inconsistent across studies. Some studies have found gender differences in health services utilization [17-22], whereas others have found no statistically significant evidence of a relationship [23,24]. The study sample was educated and age-restricted, specifically, young people and college students, who are in the upper part of the distribution in terms of resources. To some extent, education can improve the ability of individuals to increase their knowledge of modern health care, make decisions regarding their own health, and realize the benefits of using health services. There is ample evidence that education is a significant factor of service utilization [41,42]. Thus, the educated and age-restricted sample may attenuate gender disparities in health services utilization among Taiwanese college students.

Consistent with previous studies [43-45], this study found that functional eHealth literacy was negatively related to the time interval aspect. Functional literacy refers to basic reading and writing skills and the ability to understand and use health information [5]. Poor functional health literacy hinders individuals' full understanding of personal health, disease, and treatment [46]. Individuals with low functional literacy may receive ineffective care because they do not comprehend medical care directions [44], and they may have insufficient problem-solving abilities or be unlikely to change their behavior based on new information [45]. In this study, the shorter time interval indicates the higher frequency of medical use. Taiwan's National Health Insurance (NHI) system uses a third-party payer mechanism to cover medical expenses, making it extremely convenient and inexpensive for individuals to obtain care. Short or negligible waiting times stand out as hallmarks of the NHI system. However, the NHI system has also resulted in a surge in the "volume" of care provided and a sharp rise in medical expenses. Under the convenient NHI system, Taiwanese college students with inadequate functional eHealth literacy may visit physicians frequently in order to achieve the same therapeutic goal.

This study found that interactive eHealth literacy was positively related to the type, site, and time interval aspects of health services utilization. Under Taiwan's NHI system, the insured can visit any NHI-contracted hospital, clinic, pharmacy, or medical laboratory for access to health care. In recent years, the NHI Administration inaugurated a new copayment fee schedule and referral to a multitiered health care system to encourage patients to seek treatment for basic ailments at local clinics and then obtain referrals to regional hospitals or medical centers if more advanced treatment or tests are necessary. Interactive literacy involves more advanced cognitive and literacy skills that can be used to actively participate in everyday activities and to transfer the information received to one's own situation [5]. Interactive health literacy can augment one's ability to act independently, enhance motivation, and improve self-confidence [47], thus enabling Taiwanese college students to choose suitable types of and sites for health care services. It is worth noting that interactive health communication has the potential to improve the practice of medicine and the structure of health care systems but may also cause harm [48]. Studies have found that >15% of internet users said they felt overwhelmed and confused by the amount of information they found on the web. Another 54% of internet users said that the information they found led them to ask their physician a new question or to visit another physician to obtain a second opinion [49]. Thus, Taiwanese college students with high interactive literacy may visit physicians frequently in order to clarify or confirm confusing Web-based information and their illness condition.

This study found that critical eHealth literacy was positively related to the type, site, and purpose aspects of health services utilization. Taiwan's NHI system provides not only outpatient care service but also integrated preventive health care service. The goals of integrated preventive health care service include increasing the utilization of preventive care services, the early detection and intervention of high-risk and preclinical cases, and integrating the three levels of prevention, namely, health promotion and education, early detection and treatment, and the medical care system. Critical literacy involves the most advanced cognitive skills, which can be applied to critically analyze information and to use this information to exert greater control over life events and situations [5]. Critical literacy allows individuals to evaluate health issues and recognize risks and benefits as well as to advocate for themselves [50], thus enabling Taiwanese college students to choose suitable types of and sites for health care services. It also increased the utilization of health care services based on different needs (eg, preventive care, custodial care, and illness-related care).

Finally, the study found that functional eHealth literacy showed the lowest correlation to health services utilization among the three eHealth literacy levels. Functional health literacy rarely involves interactive communication and does not provide individuals with the skills to take action in their own community [51]. It is not sufficient for individuals to obtain health information; they must further evaluate and use it to make decisions about their health. Thus, in the process of functional eHealth literacy, individuals do not engage as deeply with issues as they do in the processes of interactive and critical eHealth literacy. Another explanation is that age and education play considerable roles in functional eHealth literacy. Many studies have shown that adults with functional health literacy are more likely to be younger and have higher levels of education [52,53]. The participants (average age, 21.51 years) in this study were higher education students, which facilitates functional eHealth literacy. Functional eHealth literacy showed the lowest correlation to health services utilization among the three eHealth literacy levels in this educated and age-restricted sample. Previous studies have also found that functional eHealth literacy has the lowest correlation to health behaviors and health-promoting lifestyles among college students [6-8].

Limitations

The study sample was educated and age-restricted. Population characteristics are noted as direct factors of health services utilization [11,13]. However, this study did not collect information on participant variables such as socioeconomic status or marital status. Thus, the findings should not be overgeneralized and should be interpreted in light of the sample's homogeneity. Additionally, our measure of health services utilization was not validated in the published literature. Future studies should develop and construct an HSUS. Moreover, this study found that functional eHealth literacy was related only to the time interval aspect of health services utilization. Other mediating or confounding variables may exist that should be taken into consideration such as health information seeking and increased empowerment through information seeking [35].

Conclusions

The establishment of a link between the three levels of eHealth literacy and the four aspects of health services utilization is the unique contribution of this study. Over the past decade, few attempts have been made to examine other components of eHealth literacy, such as the ability to extract and critically analyze information and use it to make decisions, which should be part of interactive and critical eHealth literacy. Notably, this study found that functional and interactive eHealth literacy were related to the time interval aspect. This finding has important implications for health care providers, who should provide patients with information leaflets to help them understand their illness and, thereby, reduce the waste of medical resources. Given that no significant gender differences in health services utilization were identified, the educated and age-restricted sample appeared to attenuate gender disparities in health services utilization. The associations among gender, age, education, and health services utilization are worth examining in further studies.

Conflicts of Interest

None.

Multimedia Appendix 1

Health Services Utilization Scale.

[PDF File (Adobe PDF File), 18KB - jmir_v20i10e283_app1.pdf]

Multimedia Appendix 2

Hierarchical multiple regression analysis of four aspects of health services utilization (Model 1).

[PDF File (Adobe PDF File), 24KB - jmir_v20i10e283_app2.pdf]

Multimedia Appendix 3

Hierarchical multiple regression analysis of four aspects of health services utilization (Model 2).

[PDF File (Adobe PDF File), 29KB - jmir_v20i10e283_app3.pdf]

References

1. Karnoe A, Kayser L. How is eHealth literacy measured and what do the measurements tell us? A systematic review. Knowledge Management & E-Learning 2015:576-600.
2. van der Varrt R, van Deursen J, Drossaert CH, Taal E, van Dijk JA, van de Laar MA. Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011;13(4):e86 [FREE Full text] [doi: 10.2196/jmir.1840] [Medline: 22071338]
3. Cutilli CC, Bennett IM. Understanding the health literacy of America: results of the National Assessment of Adult Literacy. Orthop Nurs 2009 Feb;28(1):27-34 [FREE Full text] [doi: 10.1097/01.NOR.0000345852.22122.d6] [Medline: 19190475]
4. Norman CD, Skinner HA. eHealth Literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
5. Nutbeam D. The evolving concept of health literacy. Soc Sci Med 2008 Dec;67(12):2072-2078. [doi: 10.1016/j.socscimed.2008.09.050] [Medline: 18952344]
6. Hsu WC, Chiang CH, Yang SC. The effect of individual factors on health behaviors among college students: the mediating effects of eHealth literacy. J Med Internet Res 2014;16(12):e287 [FREE Full text] [doi: 10.2196/jmir.3542] [Medline: 25499086]

7. Yang SC, Luo YF, Chiang CH. The associations among individual factors, eHealth literacy, and health-promoting lifestyles among college students. J Med Internet Res 2017 Jan 10;19(1):e15 [FREE Full text] [doi: 10.2196/jmir.5964] [Medline: 28073739]
8. Chiang CH, Yang SC, Hsu WC. Development and validation of the e-health literacy scale and investigation of the relationships between e-health literacy and healthy behavior among undergraduate students in Taiwan. Formosa J Ment Health 2015;28(3):389-420.
9. Graffigna G, Barello S, Bonanomi A, Riva G. Factors affecting patients' online health information-seeking behaviours: the role of the Patient Health Engagement (PHE) Model. Patient Educ Couns 2017 Oct;100(10):1918-1927. [doi: 10.1016/j.pec.2017.05.033] [Medline: 28583722]
10. van der Vaart R, Drossaert C. Development of the digital health literacy instrument: measuring a broad spectrum of health 1.0 and health 2.0 skills. J Med Internet Res 2017 Jan 24;19(1):e27 [FREE Full text] [doi: 10.2196/jmir.6709] [Medline: 28119275]
11. Andersen RM. Revisiting the behavioral model and access to medical care: does it matter? J Health Soc Behav 1995 Mar;36(1):1-10. [Medline: 7738325]
12. Aday LA, Andersen R. A framework for the study of access to medical care. Health Serv Res 1974;9(3):208-220 [FREE Full text] [Medline: 4436074]
13. Anderson JG. Health services utilization: framework and review. Health Serv Res 1973;8(3):184-199 [FREE Full text] [Medline: 4593850]
14. Andersen RM, Yu H, Wyn R, Davidson PL, Brown ER, Teleki S. Access to medical care for low-income persons: how do communities make a difference? Med Care Res Rev 2002 Dec;59(4):384-411. [doi: 10.1177/107755802237808] [Medline: 12508702]
15. Babitsch B, Gohl D, von Lengerke T. Re-revisiting Andersen's Behavioral Model of Health Services Use: a systematic review of studies from 1998-2011. Psychosoc Med 2012;9:Doc11 [FREE Full text] [doi: 10.3205/psm000089] [Medline: 23133505]
16. Parslow R, Jorm A, Christensen H, Jacomb P. Factors associated with young adults' obtaining general practitioner services. Aust Health Rev 2002;25(6):109-118. [Medline: 12536870]
17. Dhingra SS, Zack M, Strine T, Pearson WS, Balluz L. Determining prevalence and correlates of psychiatric treatment with Andersen's behavioral model of health services use. Psychiatr Serv 2010 May;61(5):524-528. [doi: 10.1176/ps.2010.61.5.524] [Medline: 20439377]
18. Kim HK, Lee M. Factors associated with health services utilization between the years 2010 and 2012 in Korea: using Andersen's Behavioral model. Osong Public Health Res Perspect 2016 Feb;7(1):18-25 [FREE Full text] [doi: 10.1016/j.phrp.2015.11.007] [Medline: 26981338]
19. Sareen J, Jagdeo A, Cox BJ, Clara I, ten Have M, Belik S, et al. Perceived barriers to mental health service utilization in the United States, Ontario, and the Netherlands. Psychiatr Serv 2007 Mar;58(3):357-364. [doi: 10.1176/ps.2007.58.3.357] [Medline: 17325109]
20. Vega WA, Kolody B, Aguilar-Gaxiola S, Catalano R. Gaps in service utilization by Mexican Americans with mental health problems. Am J Psychiatry 1999 Jun;156(6):928-934. [doi: 10.1176/ajp.156.6.928] [Medline: 10360134]
21. Ahmed SM, Adams AM, Chowdhury M, Bhuiya A. Gender, socioeconomic development and health-seeking behaviour in Bangladesh. Soc Sci Med 2000 Aug;51(3):361-371. [Medline: 10855923]
22. Trivedi AN, Ayanian JZ. Perceived discrimination and use of preventive health services. J Gen Intern Med 2006 Jun;21(6):553-558 [FREE Full text] [doi: 10.1111/j.1525-1497.2006.00413.x] [Medline: 16808735]
23. Lin CY. The influence of medical internet information on demographic characteristics consumer's medical behavior: a study of regional teaching hospital in Chiayi. Unpublished master thesis. Taiwan, Chiayi: Toko University; 2013.
24. Redondo-Sendino A, Guallar-Castillón P, Banegas JR, Rodríguez-Artalejo F. Gender differences in the utilization of health-care services among the older adult population of Spain. BMC Public Health 2006 Jun 16;6:155 [FREE Full text] [doi: 10.1186/1471-2458-6-155] [Medline: 16780576]
25. Elo IT. Utilization of maternal health-care services in Peru: the role of women's education. Health Transit Rev 1992 Apr;2(1):49-69. [Medline: 10148665]
26. Shaikh BT, Hatcher J. Health seeking behaviour and health service utilization in Pakistan: challenging the policy makers. J Public Health (Oxf) 2005 Mar;27(1):49-54. [doi: 10.1093/pubmed/fdh207] [Medline: 15590705]
27. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med 2011 Jul 19;155(2):97-107. [doi: 10.7326/0003-4819-155-2-201107190-00005] [Medline: 21768583]
28. Altin SV, Stock S. The impact of health literacy, patient-centered communication and shared decision-making on patients' satisfaction with care received in German primary care practices. BMC Health Serv Res 2016 Dec 30;16:450 [FREE Full text] [doi: 10.1186/s12913-016-1693-y] [Medline: 27577702]
29. Scott TL, Gazmararian JA, Williams MV, Baker DW. Health literacy and preventive health care use among Medicare enrollees in a managed care organization. Med Care 2002 May;40(5):395-404. [Medline: 11961474]

30. Ricciardi L, Mostashari F, Murphy J, Daniel JG, Siminerio EP. A national action plan to support consumer engagement via e-health. Health Aff (Millwood) 2013 Feb;32(2):376-384. [doi: 10.1377/hlthaff.2012.1216] [Medline: 23381531]
31. Neter E, Brainin E, Baron-Epel O. The third digital divide in the health domain: is Internet use for health purposes associated with health benefits? In: Hale TM, Chou WYS, Cotten SR, editors. eHealth: current evidence, promises, perils and future directions. England: Englandmerald Publishing Limited; 2018:153-175.
32. Mueller J, Jay C, Harper S, Todd C. The role of web-based health information in help-seeking behavior prior to a diagnosis of lung cancer: a mixed-methods study. J Med Internet Res 2017 Jun 08;19(6):e189 [FREE Full text] [doi: 10.2196/jmir.6336] [Medline: 28596146]
33. Mitsutake S, Shibata A, Ishii K, Oka K. Association of eHealth literacy with colorectal cancer knowledge and screening practice among internet users in Japan. J Med Internet Res 2012;14(6):e153 [FREE Full text] [doi: 10.2196/jmir.1927] [Medline: 23149453]
34. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
35. Schulz PJ, Fitzpatrick MA, Hess A, Sudbury-Riley L, Hartung U. Effects of eHealth literacy on general practitioner consultations: a mediation analysis. J Med Internet Res 2017 May 16;19(5):e166 [FREE Full text] [doi: 10.2196/jmir.6317] [Medline: 28512081]
36. Ding L, Velicer W, Harlow L. Effect of estimation methods, number of indicators per factor and improper solutions on structural equation modeling fit indices. Structural Equation Modeling 1995;2(2):119-143.
37. Gorsuch RL. Factor Analysis. Hillsdale, NJ: Lawrence Erlbaum; 1983.
38. Bentler P, Bonnett DG. Significant tests and goodness of fit in the analysis of covariance structure. Psychological Bulletin 1980;88(3):588-606.
39. Hooper D, Coughlan J, Mullen M. Structural equation modelling: guidelines for determining model fit. Electronic Journal of Business Research Methods 2008;6(1):53-60.
40. Jöreskog KG, Sörbom D. LISREL 8: Structural Equation Modeling with SIMPLIS Command Language. Chicago, IL: ScientificSoftware International; 1993.
41. Babalola S, Fatusi A. Determinants of use of maternal health services in Nigeria - looking beyond individual and household factors. BMC Pregnancy Childbirth 2009 Sep 15;9:43 [FREE Full text] [doi: 10.1186/1471-2393-9-43] [Medline: 19754941]
42. Ensor T, Cooper S. Overcoming barriers to health service access: influencing the demand side. Health Policy Plan 2004 Mar;19(2):69-79. [Medline: 14982885]
43. Cho YI, Lee SD, Arozullah AM, Crittenden KS. Effects of health literacy on health status and health service utilization amongst the elderly. Soc Sci Med 2008 Apr;66(8):1809-1816. [doi: 10.1016/j.socscimed.2008.01.003] [Medline: 18295949]
44. Baker DW, Parker RM, Williams MV, Clark WS, Nurss J. The relationship of patient reading ability to self-reported health and use of health services. Am J Public Health 1997 Jun;87(6):1027-1030. [Medline: 9224190]
45. Baker DW, Parker RM, Williams MV, Clark WS. Health literacy and the risk of hospital admission. J Gen Intern Med 1998 Dec;13(12):791-798 [FREE Full text] [Medline: 9844076]
46. Kalichman SC, Benotsch E, Suarez T, Catz S, Miller J, Rompa D. Health literacy and health-related knowledge among persons living with HIV/AIDS. Am J Prev Med 2000 May;18(4):325-331. [Medline: 10788736]
47. Sørensen K, Van den Broucke S, Fullam J, Doyle G, Pelikan J, Slonska Z, et al. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health 2012 Jan 25;12:80 [FREE Full text] [doi: 10.1186/1471-2458-12-80] [Medline: 22276600]
48. Robinson TN, Patrick K, Eng TR, Gustafson D. An evidence-based approach to interactive health communication: a challenge to medicine in the information age. Science Panel on Interactive Communication and Health. JAMA 1998 Oct 14;280(14):1264-1269. [Medline: 9786378]
49. Fox S. Online health search 2006. 2006. URL: http://goo.gl/Sq5ML4 [accessed 2018-09-19] [WebCite Cache ID 72XQeYEPS]
50. Mitchell B, Begoray D. Electronic personal health records that promote self-management in chronic illness. Online J Issues Nurs 2010;15(3):1B-10B.
51. St Leger L. Schools, health literacy and public health: possibilities and challenges. Health Promot Int 2001 Jun;16(2):197-205. [Medline: 11356758]
52. Wolf MS, Gazmararian JA, Baker DW. Health literacy and functional health status among older adults. Arch Intern Med 2005 Sep 26;165(17):1946-1952. [doi: 10.1001/archinte.165.17.1946] [Medline: 16186463]
53. Yoshida Y, Iwasa H, Kumagai S, Suzuki T, Yoshida H. Limited functional health literacy, health information sources, and health behavior among community-dwelling older adults in Japan. ISRN Geriatrics 2014;2014:1-6. [doi: 10.1155/2014/952908]

Abbreviations

AGFI: adjusted goodness of fit index
CFA: confirmatory factor analysis
CFI: comparative fit index

eHealth: electronic health
eHLS: eHealth Literacy Scale
GFI: goodness of fit index
HSUS: Health Services Utilization Scale
IFI: incremental fit index
NFI: normed fit index
NHI: National Health Insurance
PCFI: parsimonious comparative fit index
PGFI: parsimonious goodness of fit index
PNFI: parsimonious normed fit index
RFI: relative fit index
RMR: root mean square residual
RMSEA: root mean square error of approximation
SRMR: standardized root mean square residual
TLI: Tucker-Lewis index

Edited by G Eysenbach; submitted 07.09.17; peer-reviewed by E Lattie, J Nguyen, E Neter, N Stetten, L Sbaffi, R El Sherif, D Nault; comments to author 30.11.17; revised version received 19.01.18; accepted 28.07.18; published 25.10.18.

Please cite as:
Luo YF, Yang SC, Chen AS, Chiang CH
Associations of eHealth Literacy With Health Services Utilization Among College Students: Cross-Sectional Study
J Med Internet Res 2018;20(10):e283
URL: http://www.jmir.org/2018/10/e283/
doi:10.2196/jmir.8897
PMID:30361201

Original Paper

Correlation Between eHealth Literacy and Health Literacy Using the eHealth Literacy Scale and Real-Life Experiences in the Health Sector as a Proxy Measure of Functional Health Literacy: Cross-Sectional Web-Based Survey

Pietro Del Giudice[1], MD, MPH; Giulia Bravo[1], PhD; Marco Poletto[1], MD; Anna De Odorico[1], BA; Alessandro Conte[1], MD, MPH; Laura Brunelli[1], MD; Luca Arnoldo[1], MD, MPH; Silvio Brusaferro[1], MD

[1] Department of Medicine, University of Udine, Udine, Italy

Corresponding Author:
Giulia Bravo, PhD
Department of Medicine
University of Udine
Piazzale M Kolbe, 4
Udine, 33100
Italy
Phone: 39 0432559601
Email: giulia.bravo@uniud.it

Abstract

Background: The eHealth Literacy Scale (eHEALS) is a tool for the self-assessment of perceived comfort and skills in using the internet as a source for health-related information. Although evidence exists of the reliability and construct and structural validity of the scale, there is a lack of evidence in relation to what is proposed by Norman and Skinner in their theoretical lily model of eHealth literacy; in particular it is not clear whether having a higher level of health literacy can positively influence electronic health (eHealth) literacy as measured by the eHEALS.

Objective: Our study aim was to assess whether real-life experiences from studying or working in the health field, as a proxy of higher functional health literacy, correlate with self-referred eHealth literacy as measured by the eHEALS.

Methods: A Web-based survey was conducted among adults living in Northeast Italy using an Italian version of the eHEALS (IT-eHEALS). In order to be able to measure the effect of higher functional health literacy on eHealth literacy, we divided our sample into two groups, respectively characterized by studying or working experience in the health sector and by lack thereof. Mean differences between eHEALS were calculated using t test and effect size evaluated using Cohen d. To ensure the validity of the IT-eHEALS, we evaluated its psychometric properties (internal consistency and dimensionality) and construct validity (by evaluating its correlation with respondents age, gender, educational attainment, self-rated health, use of internet for health-related purposes, and working status).

Results: A total of 868 respondents that completed the IT-eHEALS were included for analysis, of which 259 had working or studying experience in the health field. Mean (SD) eHEALS total score was 28.2 (6.2) for the whole sample, with statistically significant differences (*P*<.001) between the two groups, with the higher health literate group scoring significantly better (31.9 (5.9) vs 26.7 (5.6), respectively), with a standardized mean difference (Cohen d) of 0.9. Interestingly, we found a weak, yet significant, correlation between eHealth literacy and respondent characteristics for the higher health literate group only, as measured by positive Spearman correlation coefficients for age (0.11, *P*=.001), educational attainment (0.19, *P*=.002) and self-rated health (0.14, *P*=.024). Also, in line with current literature, correlation of eHEALS score with frequency of internet use for health-related purposes was significant for both groups (0.32, *P*<.001 and 0.15, *P*<.001 for higher and lower health literacy group, respectively). In our study we could not find any difference related to gender, while a significant difference for working status was only present when considering the sample as a whole (*P*=.03).

Conclusions: Our study demonstrates a sizeable effect of higher levels of functional health literacy on the eHEALS score, corroborating what was initially proposed by Norman and Skinner in the lily model of eHealth literacy.

(J Med Internet Res 2018;20(10):e281) doi:10.2196/jmir.9401

KEYWORDS

eHealth literacy; health literacy; effect size; eHEALS; lily model

Introduction

Health Information and the Internet

Use of the internet for health-related purposes poses a particularly important challenge, as it has been shown that wrong or incomplete information available on the internet may have negative consequences for the user—including on doctor-patient relationships, participation in prevention and screening programs, or adherence to medical treatment [1]. Today, the availability and accessibility of quality health-related internet information is still an issue, and agreement on a specified set of quality standards for health websites has recently been proposed as a new public health priority [2]. The problem of providing quality health-related information has become even more complex in the current Web 2.0 environment, as the search strategy for relevant information depends not only on the searcher's ability, but also on the influence of intermediators and apomediators, with the latter effectively pushing the search towards or away from relevant items [3].

eHealth Literacy and the eHEALS

In 2006, after three years of experimentation in a teenage health promotion program, Norman and Skinner developed the concept of eHealth literacy, drawing from the increasingly popular concept of health literacy. In the same year, the authors proposed both a theoretical model [4] and a tool to measure the new construct [5]. In their view, eHealth literacy was defined as "the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem." The proposed model, called "lily model", described eHealth literacy as the interplay of six core skills or literacies (traditional literacy, health literacy, information literacy, scientific literacy, media literacy, and computer literacy). The measurement tool, called eHealth literacy scale (eHEALS), was based on the principles of the social cognitive theory and self-efficacy theory, using Likert scales for self-assessed, subjective responses, so that measures should be considered as precursors of behavior change and skill development [6]. In this sense the eHEALS can be considered a measure of subjective, self-assessed eHealth literacy.

To further improve this first attempt to measure eHealth literacy, recent years have seen the development and validation of more comprehensive, and thus complex, eHealth literacy evaluation tools [7-9]. It should be noted, though, that despite all the possible issues coming from the simple eight-item structure of the eHEALS, its simplicity is also a strength of the tool. In fact, at the time of its development, the eHEALS was explicitly conceived so that it would have been easy to administer, taking into consideration the expressed needs of health professionals who said that they wouldn't use a long instrument in their practice [10]. This "strength in simplicity" facilitated the adoption of the eHEALS, and its widespread use has been highlighted in the findings of several literature reviews. In 2012 Collins and colleagues reviewed the use of health literacy screening tools in eHealth applications [11] and found that the eHEALS was the most used for the purpose of developing a computer-based instrument to screen individuals accessing eHealth applications, alone or in combination with other screening tools for health literacy. In a 2015 review of existing tools to measure eHealth literacy and their use by Karnoe and Kayser [12], the authors found that, out of eight different tools for measuring eHealth literacy, only the eHEALS had been used in studies other than the one it was originally published in. The same authors argue that the eHEALS, while easy to administer, provides a measure that is not able to pinpoint whether inadequate eHealth literacy is a result of insufficient health literacy, digital literacy, or a combination hereof. In other words, it is still not clear whether a higher level of self-referred eHealth literacy (using the eHEALS), is correctly due to differences in levels of functional health literacy or is just a result of high levels of perceived self-efficacy.

Aim of Current Study

The sheer amount of unchecked health-related information on the internet can be seen either as a limit or as a resource by different respondents with different skills and experiences in the health field. Considering the subjective, self-referred nature of the eHEALS, one possibility could be that people with less knowledge in the health field would trust the information more as they would be less able to discern the real quality of their internet search findings, scoring higher in the scale. The aim of our study is to test the lily model, by assessing whether and to what extent differences in health literacy levels account for variations in the eHEALS score. To our knowledge, no prior study using the eHEALS explored whether the scale behaves as intended in populations with different sets of core skills or literacies as described in the lily model, in our case based on differences in health literacy levels. In their description of the lily model of eHealth literacy, Norman and Skinner use the definition of health literacy given by the American Medical Association [13], which can be arguably referred to as a basic "functional" level of health literacy [14]. There are several possible measures of functional health literacy using measurement tools that are grounded on different theories. Yet, functional measures of health literacy have been shown to correlate with studying or working experiences in the health field [15], with physicians, researcher in the health field, and nurses showing higher health literacy levels in comparison with the general population. Therefore, we chose to recruit a large sample of respondents, divided into two groups, by asking them whether they had real-life experiences in studying or working in the health sector (eg, physicians, nurses, health alliance professionals). By doing so, we were able to compare a highly health-literate group with the rest of the general population.

Methods

Survey Design and Administration

In order to test our hypothesis, during November and December 2016 a Web-based survey was conducted by contacting people using two different recruitment strategies. Recruitment was performed using: (a) the mailing list of the student body (undergraduate and post-graduate) from the University of Udine (obtained with permission from the University), and (b) Facebook contacts of the public health research team members, who were then asked via Facebook to further disseminate the survey to their contacts. Decision to participate in the survey was voluntary and no incentives were offered to respondents. The survey was first pretested for usability and functionality by the members of the research team. The survey was administered using the software SurveyMonkey. All participants were asked to read and approve an informed consent form telling them that the study was managed by the University of Udine and that the survey would require approximately 15 minutes. As the survey did not collect any data that could be directly linked to participants' sensitive data or information that could potentially affect their health, no approval by the Ethical Committee was deemed required under Italian legislation.

Measures

Collected measures covered socio-demographic characteristics (gender, age, highest educational level attained, working status), self-perceived health status, internet health-related behaviors (use for health-related search and frequency), working or studying experiences in the health sector, and an Italian adaptation of the eHEALS scale. Age was collected as a discrete variable, in number of years. Educational attainment was first collected using an 8-item scale, later aggregated into a 3-item scale in line with the aggregation methodology used by Eurostat in relation to International Standard Classification of Education levels [16]. The final set of education levels used for analysis were: (low) 8th grade or lower, (middle) 9-13th grade, (high) university degree or higher. Working status was collected asking participants whether they were currently working, studying, or neither working nor studying (classified as "other"). Self-rated health was collected using a 5-item Likert scale, ranging from "very bad" to "excellent," with the midpoint rated as "good." Health-related internet use was measured asking the frequency of internet use for health-related purposes (using a 5-item Likert scale ranging from "not more than 5-6 times a year" to "several times a week"). To differentiate for real-life experiences in the health sector, participants were asked whether they had experiences in studying or working in the health sector using a yes/no question. Regarding the Italian version of the eHEALS, we were unable to retrieve a previously reported version of the tool (I-eHEALS) presented in a conference abstract by De Caro et al [17] (via request to the corresponding author), and a new Italian translation of the 8 eHEALS items was produced by the research team (IT-eHEALS). The translation process was carried out following established good practices [18]: the original English tool was initially distributed among the research team, producing a first set of translations that were later merged into a single draft version. The draft of the Italian instrument was then retranslated into English by an interpreter and reviewed by the research team for correctness. Translated items were pretested for comprehensibility on a small sample of Italian adults (N=24) and items were adjusted accordingly. Like the original version of the test, the IT-eHEALS is composed of 8 items measured with a 5-point Likert scale. For every respondent of the sample that completed all 8 IT-eHEALS items, the total score ranges from 8 to 40 (calculated by adding up the single items' scores), with a higher score indicating a higher self-referred eHealth literacy.

Statistical Analysis

Sample Selection and Descriptive Analysis

To test our hypothesis, we selected the subsample of respondents who completed all of the 8 IT-eHEALS items. All collected data were screened to search for missing values or for any incorrect data inclusion. When not plausible, records were excluded from the analysis upon discussion among the research team. Then, the sample was divided into two groups based on having experiences of studying or working in the health sector or not. In this paper, we will refer to the group currently studying or working in the health sector as EHS+, and to the other as EHS-. Descriptive statistics (frequency, percentage, mean [SD]) were calculated for socio-demographic variables (gender, age, educational attainment, and working status), self-rated health, and internet health-related behaviors for all groups. A comparative analysis using Wilcoxon-Mann Whitney test and Chi-square (or Fisher Exact) Test, respectively for continuous and categorical variables, was conducted to detect statistically significant group differences ($P<.05$).

IT-eHEALS Scale Validity

Since we used a newly developed and adapted Italian version of the eHEALS (IT-eHEALS), we also assessed the scale by examining its psychometric properties and construct validity. Psychometric properties were examined by measuring internal consistency (Cronbach alpha) and conducting a principal component analysis to assess the dimensionality of the scale. Construct validity was assessed using a hypothesis testing approach. Based on prior studies, it was hypothesized that participants who (a) are younger [19], (b) use the internet for health-related purpose more frequently [20], (c) have a better self-rated health [17,21], and (d) have higher educational attainment [19], would have higher self-referred eHealth literacy scores. Spearman rho index was used to assess correlations between IT-eHEALS total score and (a) age, (b) internet use, (c) self-rated health and (d) educational level in the two groups of IT-eHEALS respondents. Also, we used *t* test and analysis of variance (ANOVA) to evaluate the difference in IT-eHEALS scores for gender and working status, respectively.

Relation Between Health and eHealth literacy

Finally, differences between eHEALS means and SDs in the EHS+ and EHS- groups were calculated using *t* test, and effect size was evaluated using Cohen *d*. Analysis was conducted using SAS software version 9.4 for Windows (SAS Institute Inc, Cary, NC, USA).

Results

Socio-Demographic Characteristics

In total, the two internet surveys led to the recruitment of 1136, of which 868 completed all eight IT-eHEALS items, leading to a final sample of 868 respondents that were included for analysis. Table 1 shows the socio-demographic characteristics of the whole sample and differences between the EHS+ and

EHS- groups. The two groups differ significantly (P<.001) in relation to working status and frequency of internet use for health-related purposes. In the EHS+ group, most of the respondents are working (139/259, 53.7%), while in EHS- the majority are studying (303/609, 49.7%). Regarding internet use for health-related purposes, 27.4% (71/259) of EHS+ respondents use the internet more than once a week, while only 5.42% (33/609) of the EHS- respondents do so, suggesting differences in health-related internet behaviors between the two groups. Also, respondents in the EHS+ group are significantly older, with a mean age of 31.5±12.1 years vs 28.7±9.7 years for the EHS- group (*P*=.008).

Validity of the IT-eHEALS Scale

IT-eHEALS showed a high degree of internal consistency with a Cronbach alpha of .90, with slight, negligible differences between the two groups (.87 in EHS-, .91 in EHS+). Principal Component Analysis in the whole sample confirmed the unidimensionality of the scale (eigenvalue=4.9 with 61.1% of variance explained). All IT-eHEALS items show high loadings on the first component (ranging from 0.68 to 0.83). Table 2 shows Spearman correlation coefficients with age and educational attainment. Correlation coefficients of total mean scores on the IT-eHEALS with selected variables are significant but low, with the exception of age, educational attainment, and self-rated health in EHS-. The correlation with frequency of internet use for health-related purposes was significant in both groups.

We did not find any difference in relation to gender. When assessing the whole sample, there was a significant difference for working status (*P*=.03) that was not present when considering EHS+ and EHS- separately in both groups.

Table 1. Descriptive and comparative analysis of study sample.

Variable	Whole sample (N=868), n (%)	EHS+[a] (N=259), n (%)	EHS-[b] (N=609), n (%)	*P* value[c]
Gender				. 85
Male	231 (26.6)	70 (27.0)	161 (26.4)	
Female	637 (73.4)	189 (73.0)	448 (73.6)	
Educational attainment				.057
Low	22 (2.5)	5 (1.9)	17 (2.8)	
Middle	457 (52.7)	121 (46.7)	336 (55.2)	
High	383 (44.1)	129 (49.8)	254 (41.7)	
No response	6 (0.7)	4 (1.6)	2 (0.3)	
Working status				<.001
Working	391 (45.1)	139 (53.7)	252 (41.4)	
Studying	416 (47.1)	113 (43.6)	303 (49.7)	
Other	61 (7.0)	7 (2.7)	54 (8.9)	
Self-rated health				.27
Very bad	6 (0.7)	1 (0.4)	5 (0.8)	
Poor	62 (7.1)	21 (8.1)	41 (6.7)	
Good	455 (52.4)	123 (47.5)	332 (54.5)	
Very good	281 (32.4)	90 (34.7)	191 (31.4)	
Excellent	64 (7.4)	24 (9.3)	40 (6.6)	
Frequency of internet use for health-related purposes				<.001
No more than 5-6 times/year	282 (32.5)	62 (23.9)	220 (36.1)	
No more than 2-3 times/year	135 (15.5)	31 (12.0)	104 (17.1)	
Once a month	238 (27.4)	58 (22.4)	180 (29.6)	
Once a week	109 (12.6)	37 (14.3)	72 (11.8)	
Several times a week	104 (12.0)	71 (27.4)	33 (5.4)	

[a] EHS+: Group with studying or working experiences in the health sector.

[b] EHS-: Group without studying or working experiences in the health sector.

[c] *P* values are calculated for mean differences between groups EHS+ and EHS-.

Table 2. Spearman correlations between eHealth Literacy Scale total score for selected variables.

Variable	Whole sample		EHS+[a]		EHS-[b]	
	Spearman correlation coefficient	*P* value	Spearman correlation coefficient	*P* value	Spearman correlation coefficient	*P* value
Age	0.11	.002	0.22	.001	0.02	.65
Educational attainment	0.11	.001	0.19	.002	0.06	.13
Self-rated health	0.07	.038	0.14	.024	0.02	.70
Frequency of internet use for health	0.28	<.001	0.32	<.001	0.15	<.001

[a] EHS+: Group with studying or working experiences in the health sector.

[b] EHS-: Group without studying or working experiences in the health sector.

Table 3. Italian version of eHealth Literacy Scale (eHEALS) items and total score statistics.

eHEALS score	Whole sample (N=868), mean (SD)	EHS+[a] (N=259), mean (SD)	EHS-[b] (N=609), mean (SD)	*P* value[c]
Item 1	3.8 (0.9)	4.2 (0.8)	3.6 (0.8)	<.001
Item 2	3.5 (0.9)	3.9 (1.0)	3.4 (0.9)	<.001
Item 3	3.6 (1.0)	4.0 (0.9)	3.4 (0.9)	<.001
Item 4	3.7 (0.9)	4.1 (0.9)	3.5 (0.9)	<.001
Item 5	3.7 (0.9)	4.1 (0.9)	3.6 (0.9)	<.001
Item 6	3.5 (1.2)	4.2 (1.0)	3.2 (1.1)	<.001
Item 7	3.8 (1.0)	4.2 (0.8)	3.6 (1.0)	<.001
Item 8	2.7 (1.2)	3.2 (1.2)	2.4 (1.1)	<.001

[a] EHS+: Group with studying or working experiences in the health sector.

[b] EHS-: Group without studying or working experiences in the health sector.

[c] *P* values are calculated for mean differences between groups EHS+ and EHS-.

Health Literacy and the eHEALS

Table 3 shows the mean (SD) item score and the statistical significance of the difference between the EHS+ and EHS- groups (see Multimedia Appendix 1 for item descriptions). Considering the whole sample of IT-eHEALS respondents, mean values for items range from 3.8 (item 1) to 2.7 (item 8). Differences between the two groups were significant for all IT-eHEALS items (*P*<.001), with the mean (SD) total score significantly higher in EHS+ compared to EHS- (31.9 [5.9] vs 26.7 [5.6], *P*<.001). The standardized mean difference (Cohen *d*) was 0.9, demonstrating a sizeable effect of higher levels of functional health literacy on the eHEALS score.

Discussion

Study Findings

Correlation Between Health Literacy and eHealth Literacy

In our study we were able to demonstrate that real-life working or studying experiences in the health sector, as a proxy of higher levels of health literacy, positively correlate with self-referred eHealth literacy as measured by the eHEALS. This finding is in line with the original lily model of eHealth literacy proposed by Norman and Skinner, where eHealth literacy is described as the interconnection of different core skills, including health literacy. Our findings emphasize that there are different factors other than internet and computer skills that can lead to different results when measuring eHealth literacy.

Psychometric Characteristics and Construct Validity of the eHEALS

Regarding the validity of the IT-eHEALS in the Italian population, we found high internal consistency, as shown by the Cronbach alpha and the inter-item correlation analysis, with comparable results with other translation of the eHEALS [19-26]. Our principal component analysis shows that the IT-eHEALS can be better explained by a single component structure, supporting its unidimensionality. While authors of two past studies using the eHEALS argued that the scale could have been multidimensional [24,27], our results are in line with other studies that confirmed the unidimensional nature of the scale, which allow for the calculation of a total mean score of all the eHEALS items [20,28,29]. Regarding the construct validity of the eHEALS, interpretation of our findings should be taken cautiously due to possible bias introduced by the sampling technique and keeping in mind that the sample was composed of young adults aged 20-30. Also, as already noted by Diviani et al [20], most of eHEALS validation studies have been conducted among specific populations, with different results showing no consistent association of eHEALS scores with the personal characteristics of the respondents, such as gender, education, or age. In fact, while some studies found significant correlation of eHealth literacy levels with age

[19,26], education [19], gender [26] and self-rated health [17,21], other studies found no correlation for the same variables. In particular, several other studies found no correlation between the eHEALS and gender [19-21], age [20], and education [20]. Our study findings show that the IT-eHEALS have a weak, positive correlation with age, educational attainment, and self-rated health. It must be noted that, interestingly, when considering our two subsamples separately, these correlations show a level of significance only in EHS+, while this is not true for EHS-, suggesting a correlation between different levels of functional health literacy and self-referred eHealth literacy. Regarding gender, we found no correlation with the eHEALS score, a result that is comparable with other studies involving a similar young and highly educated population [20,21]. Also, in line with similar studies [20], we found a weak level of correlation with the frequency of internet use for health-related purposes in all groups. Overall, these results suggest that the eHEALS should be considered a valid tool that can be used to assess the perceived comfort and skills in using the internet for health-related purposes.

Study Limitations

Our study has some limitations that should be acknowledged.

Sample Composition

A first limitation of our study lies in the recruitment strategy used, which led to a study sample which is composed by respondents who are mostly young and highly educated, and therefore could not be considered representative of the adult Italian population, limiting the generalizability of our findings. While the English version of the scale has been applied in a variety of samples, most of the validating studies of the eHEALS in other languages have only been conducted among specific populations. Regarding gender, our sample has an overrepresentation of female respondents, so that our results shall be taken cautiously when trying to generalize to the general adult population. Also, it should be noted that the use of Facebook in our recruitment strategy made it impossible to assess number and characteristic of nonrespondents, an important limitation that should also be considered when interpreting results. While these are common shortcoming of similar validation studies, we believe that its composition characteristics (higher education level, younger age) are somewhat representative of the most active population of health information seekers in the internet, as reported by the latest 2017 EU Digital Scoreboard statistics for Italy about health information seeking in the general population (see Multimedia Appendix 2). Moreover, our study population was sufficient to address our aim, namely the recruitment of a sample large enough to be divided into two comparable groups characterized by study or work experiences in the health sector. In relation to this point, we are also aware that the two groups were not equally distributed for some of the socio-demographic and health-related internet behavior factors; since our methodology did not allow us to select our sample composition beforehand, we cannot be sure whether group differences are an effect of the selected variable for group inclusion (in our case people with experiences in the health sector having different baseline characteristics compared to the general population for age, working status and use of internet for health-related purposes), or whether there are other reasons for these differences that are not due to the recruitment techniques we used.

Measures

Another limitation of our study lies in the fact that we only included one measure of internet health-related behavior, as comparing different measures was outside the original scope of the study. While it should be acknowledged that this measure has not been previously validated, our results suggest that the two groups may indeed be different in terms of internet health-related behaviors, yet these should be further explored with a larger number of measures before reaching definitive conclusions on the health literacy role in explaining behavioral differences in this field. Also, we did not include any validated measures of either subjective or objective health literacy, which could have been used to quantitatively assess different levels of health literacy. Instead, we asked for real-life experiences in the health field as a proxy, which have been showed to correlate only with objective health literacy tests [15]. Our results show that there is a correlation between these experiences and the eHEALS, yet we suggest that future studies also include other validated measures of health literacy to better correct results and to explore the correlation of the eHEALS with both subjective and objective measures of health literacy. Another limitation of the present study is the lack of an objective measure of eHealth literacy skill, making it unclear whether the differences between groups in health-related internet behaviors could also be related to actual, objective eHealth literacy skills. At the moment, there are mixed results regarding the correlation between eHEALS and objective measures of competencies on health-related internet use: using different measures of eHealth literacy objective competencies, Neter and Brainin found moderate correlation [30], while van der Vaart et al found no correlation [23]. This is also common to other measures used in the field of health literacy and is probably due to the subjective nature of the tools used [31], and even in the presence of a moderate correlation, Neter and Brainin recommend assessing the two constructs separately [30]. As these authors are providing the methodological base for more objective eHealth literacy measures, we also encourage future studies to include measures based on these methodologies [30,32]. This would not only lead to a better comprehension of the relation between subjective and objective measures, but it would also contribute to the possibility of expanding the item bank of objective measures for future studies, with possible use of advanced theories for test development such as Item Response Theory or the Rasch model [20,31].

eHEALS Version

It must be noted that after our study was conducted, a validation study of another Italian version of the eHEALS (I-eHEALS) was published by Diviani et al, using a sample population of Italian-speaking Swiss respondents [20]. As we used a different Italian translation of the eHEALS, it remains unclear whether results could be comparable to their results. While there are minor differences in the phrasing of the items, our scale shows good internal consistency and construct validity. For this reason, we believe that the two currently available Italian translations of the eHEALS (I-eHEALS and IT-eHEALS) can both be considered valid and, in our opinion, can be used interchangeably (see Multimedia Appendix 1 for eHealth Literacy Scale Italian versions).

Conclusions

This study demonstrates that, as proposed in the lily model of eHealth literacy, eHEALS scale results are affected by a higher level of health literacy, measured via real-life experiences in the field of health as a proxy. We believe that this is an original result, which could be relevant in the current stage of scientific discussion regarding the use of the eHEALS and further advancements in measuring eHealth literacy. Despite its several limitations, and in absence of simple, easy-to-administer measurement tools, the eHEALS can still be considered a valid tool to assess self-perceived comfort and skills in using the internet for health-related purposes. It should still be used for comparison in the elaboration of new eHealth literacy measures, which should be designed including new items and different subscales in order to be able to capture all the proposed “literacies” of the construct [4]. For these reasons, we believe that the absence of correlation of the eHEALS with objectively measured internet related skills as found by different authors does not undermine the validity and the usability of the scale per se, and that the eHEALS can still be applied in clinical and health promotion activities, for example to identify different needs for the participants to an eHealth intervention or to evaluate intervention results.

Authors' Contributions

PDG, MP, ADO, LB, LA, and AC discussed and drafted the questionnaire for the surveys and discussed and approved the IT-eHEALS translation of the original eHEALS. AC and ADO managed data collection activities. PDG and GB analyzed collected data and interpreted results. PDG and GB drafted the manuscript, which was revised and approved by MP and SB. The study did not receive any funding.

Conflicts of Interest

None declared.

Multimedia Appendix 1

eHealth Literacy Scale Italian versions.

[PDF File (Adobe PDF File), 140KB - jmir_v20i10e281_app1.pdf]

Multimedia Appendix 2

Internet health-information seeking behavior in the Italian adult population.

[PDF File (Adobe PDF File), 349KB - jmir_v20i10e281_app2.pdf]

References

1. Cline RJ, Haynes KM. Consumer health information seeking on the Internet: the state of the art. Health Educ Res 2001 Dec;16(6):671-692 [FREE Full text] [Medline: 11780707]
2. Devine T, Broderick J, Harris LM, Wu H, Hilfiker SW. Making Quality Health Websites a National Public Health Priority: Toward Quality Standards. J Med Internet Res 2016;18(8):e211 [FREE Full text] [doi: 10.2196/jmir.5999] [Medline: 27485512]
3. Eysenbach G. Medicine 2.0: social networking, collaboration, participation, apomediation, and openness. J Med Internet Res 2008;10(3):e22 [FREE Full text] [doi: 10.2196/jmir.1030] [Medline: 18725354]
4. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. J Med Internet Res 2006 Jun;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
5. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
6. Bandura A. Self-Efficacy: The Exercise of Control. New York: W.H. Freeman and Company; 1997.
7. Kelly L, Ziebland S, Jenkinson C. Measuring the effects of online health information: Scale validation for the e-Health Impact Questionnaire. Patient Educ Couns 2015 Nov;98(11):1418-1424 [FREE Full text] [doi: 10.1016/j.pec.2015.06.008] [Medline: 26162953]
8. Britt RK, Hatten KN. The Development and Validation of the eHealth Competency Scale: A Measurement of Self-Efficacy, Knowledge, Usage, and Motivation. Technical Communication Quarterly 2016 Apr 07;25(2):137-150. [doi: 10.1080/10572252.2016.1149621]
9. Koopman RJ, Petroski GF, Canfield SM, Stuppy JA, Mehr DR. Development of the PRE-HIT instrument: patient readiness to engage in health information technology. BMC Fam Pract 2014;15:18 [FREE Full text] [doi: 10.1186/1471-2296-15-18] [Medline: 24472182]

10. Norman C. eHealth literacy 2.0: problems and opportunities with an evolving concept. J Med Internet Res 2011 Dec;13(4):e125 [FREE Full text] [doi: 10.2196/jmir.2035] [Medline: 22193243]
11. Collins SA, Currie LM, Bakken S, Vawdrey DK, Stone PW. Health literacy screening instruments for eHealth applications: a systematic review. J Biomed Inform 2012 Jun;45(3):598-607 [FREE Full text] [doi: 10.1016/j.jbi.2012.04.001] [Medline: 22521719]
12. Karnoe A, Kayser L. How is eHealth literacy measured and what do the measurements tell us? A systematic review. Knowledge Management & E-Learning 2015 Dec;7(4):576-600 [FREE Full text]
13. Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs, American Medical Association. Health literacy: report of the Council on Scientific Affairs. JAMA 1999 Feb 10;281(6):552-557. [Medline: 10022112]
14. Nutbeam D. Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. Health Promotion International 2000 Sep 01;15(3):259-267 [FREE Full text] [doi: 10.1093/heapro/15.3.259]
15. Paiva D, Silva S, Severo M, Ferreira P, Santos O, Lunet N, et al. Cross-cultural adaptation and validation of the health literacy assessment tool METER in the Portuguese adult population. Patient Educ Couns 2014 Nov;97(2):269-275. [doi: 10.1016/j.pec.2014.07.024] [Medline: 25107513]
16. Eurostat. International Standard Classification of Education (ISCED) URL: http://ec.europa.eu/eurostat/statistics-explained/index.php/International_Standard_Classification_of_Education_%28ISCED%29 [accessed 2018-06-20] [WebCite Cache ID 70JugcUjL]
17. De Caro W, Corvo E, Marucci AR, Mitello L, Lancia L, Sansoni J. eHealth Literacy Scale: An Nursing Analisys and Italian Validation. Stud Health Technol Inform 2016;225:949. [Medline: 27332425]
18. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, ISPOR Task Force for TranslationCultural Adaptation. Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health 2005;8(2):94-104. [doi: 10.1111/j.1524-4733.2005.04054.x] [Medline: 15804318]
19. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
20. Diviani N, Dima AL, Schulz PJ. A Psychometric Analysis of the Italian Version of the eHealth Literacy Scale Using Item Response and Classical Test Theory Methods. J Med Internet Res 2017 Apr 11;19(4):e114 [FREE Full text] [doi: 10.2196/jmir.6749] [Medline: 28400356]
21. Paramio Pérez G, Almagro BJ, Hernando Gómez A, Aguaded Gómez JI. [Validation of the eHealth Literacy Scale (eHEALS) in Spanish University Students]. Rev Esp Salud Publica 2015;89(3):329-338 [FREE Full text] [doi: 10.4321/S1135-57272015000300010] [Medline: 26388346]
22. Koo M, Norman C, Chang H. Psychometric evaluation of a chinese version of the eHealth literacy scale (eHEALS) in school age children. International Electronic Journal of Health Education 2012;15:29-36.
23. van der Vaart R, van Deursen AJ, Drossaert CH, Taal E, van Dijk JA, van de Laar MA. Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011;13(4):e86 [FREE Full text] [doi: 10.2196/jmir.1840] [Medline: 22071338]
24. Soellner R, Huber S, Reder M. The Concept of eHealth Literacy and Its Measurement. Journal of Media Psychology 2014 Jan;26(1):29-38. [doi: 10.1027/1864-1105/a000104]
25. Bazm S, Mirzaei M, Fallahzadeh H, Bazm R. Validity and Reliability of the Iranian Version of eHealth Literacy Scale. Journals of Community Health Research. Journals of Community Health Research; 2016;5(2):121-130 [FREE Full text]
26. Mitsutake S, Shibata A, Ishii K, Okazaki K, Oka K. [Developing Japanese version of the eHealth Literacy Scale (eHEALS)]. Nihon Koshu Eisei Zasshi 2011 May;58(5):361-371. [Medline: 21905612]
27. Stellefson M, Paige SR, Tennant B, Alber JM, Chaney BH, Chaney D, et al. Reliability and Validity of the Telephone-Based eHealth Literacy Scale Among Older Adults: Cross-Sectional Survey. J Med Internet Res 2017 Oct 26;19(10):e362 [FREE Full text] [doi: 10.2196/jmir.8481] [Medline: 29074471]
28. Nguyen J, Moorhouse M, Curbow B, Christie J, Walsh-Childers K, Islam S. Construct Validity of the eHealth Literacy Scale (eHEALS) Among Two Adult Populations: A Rasch Analysis. JMIR Public Health Surveill 2016 May;2(1):e24 [FREE Full text] [doi: 10.2196/publichealth.4967] [Medline: 27244771]
29. Richtering SS, Morris R, Soh S, Barker A, Bampi F, Neubeck L, et al. Examination of an eHealth literacy scale and a health literacy scale in a population with moderate to high cardiovascular risk: Rasch analyses. PLoS One 2017;12(4):e0175372 [FREE Full text] [doi: 10.1371/journal.pone.0175372] [Medline: 28448497]
30. Neter E, Brainin E. Perceived and Performed eHealth Literacy: Survey and Simulated Performance Test. JMIR Hum Factors 2017 Jan 17;4(1):e2 [FREE Full text] [doi: 10.2196/humanfactors.6523] [Medline: 28096068]
31. Nguyen TH, Paasche-Orlow MK, McCormack LA. The State of the Science of Health Literacy Measurement. Stud Health Technol Inform 2017;240:17-33. [Medline: 28972507]
32. van der Vaart R, Drossaert C. Development of the Digital Health Literacy Instrument: Measuring a Broad Spectrum of Health 1.0 and Health 2.0 Skills. J Med Internet Res 2017 Jan 24;19(1):e27 [FREE Full text] [doi: 10.2196/jmir.6709] [Medline: 28119275]

Abbreviations

ANOVA: analysis of variance
eHEALS: eHealth Literacy Scale
eHealth: electornic health
EHS+: Group with studying or working experiences in the health sector
EHS-: Group without studying or working experiences in the health sector
I-eHEALS: Swiss-Italian version of the eHealth Literacy scale
IT-eHEALS: Italian version of eHealth Literacy Scale

Edited by G Eysenbach; submitted 29.12.17; peer-reviewed by E Neter, L Kelly; comments to author 15.03.18; revised version received 28.06.18; accepted 20.07.18; published 31.10.18.

Please cite as:
Del Giudice P, Bravo G, Poletto M, De Odorico A, Conte A, Brunelli L, Arnoldo L, Brusaferro S
Correlation Between eHealth Literacy and Health Literacy Using the eHealth Literacy Scale and Real-Life Experiences in the Health Sector as a Proxy Measure of Functional Health Literacy: Cross-Sectional Web-Based Survey
J Med Internet Res 2018;20(10):e281
URL: http://www.jmir.org/2018/10/e281/
doi:10.2196/jmir.9401
PMID:30381283

Original Paper

Investigating the Perceptions of Primary Care Dietitians on the Potential for Information Technology in the Workplace: Qualitative Study

Aimee Jones[1], BSc (Hons), APD; Lana J Mitchell[1,2], AdvAPD, PhD; Rochelle O'Connor[1], BSc (Hons), APD; Megan E Rollo[3], PhD, APD; Katherine Slater[1], BSc (Hons), APD; Lauren T Williams[1,2], PhD, FDAA; Lauren Ball[1,2], PhD

[1]School of Allied Health Sciences, Griffith University, Gold Coast, Australia
[2]Menzies Health Institute Queensland, Griffith University, Gold Coast, Australia
[3]Priority Research Centre for Physical Activity and Nutrition, University of Newcastle, Newcastle, Australia

Corresponding Author:
Lauren Ball, PhD
School of Allied Health Sciences
Griffith University
G01 2.05A
Gold Coast,
Australia
Phone: 61 0413031470
Email: l.ball@griffith.edu.au

Abstract

Background: Chronic diseases are the leading cause of morbidity and mortality worldwide. The primary health care setting is an effective avenue for the management and prevention of chronic diseases. Dietitians working in this setting assist with the management of modifiable risk factors of chronic diseases. However, health care professionals report challenges in providing care in this setting because of time and financial constraints. Information technology offers the potential to improve health care quality, safety, efficiency, and cost-efficiency, but there exists limited understanding of dietitians' application of technology in this setting.

Objective: The objective of this study was to explore the perceptions of primary care dietitians about using information technology in their workplace.

Methods: We recruited 20 Australian primary care dietitians using purposive and snowball sampling for semistructured telephonic interviews. Interview questions aimed to gain an understanding of dietitians' perceptions about sharing patient outcomes through a national database and the benefits, disadvantages, feasibility, and barriers of using information technology. Interviews were audiorecorded, transcribed verbatim, and thematically analyzed for emerging themes and subthemes. Finally, the technologies used by participants were collated by name and researched for their key attributes.

Results: The following 4 distinct themes emerged from the data: information technology improving the efficiency of practice tasks, experiencing barriers to using information technology in practice, information technology enhancing outcomes through education and monitoring, and information technology for sharing information with others. Participants identified several advantages and disadvantages of using technology and expressed willingness to share patient outcomes using a Web-based database.

Conclusions: This study suggests that information technology is perceived to have benefits to dietitians and patients in primary health care. However, to achieve the optimal benefit, support is required to overcome barriers to integrate information technology into practice better. Further development of patient management systems and standardized Web-based data collection systems are needed to support better usage by dietitians.

(J Med Internet Res 2018;20(10):e265) doi:10.2196/jmir.9568

KEYWORDS

dietetics; information technology; mobile phone; primary health care; private practice

Introduction

Chronic diseases are a leading cause of morbidity and mortality worldwide [1]. The occurrence of chronic diseases is associated with preventable risk factors such as high blood pressure, high blood cholesterol, and overweight or obesity [1]. Dietitians are uniquely qualified to apply the science of nutrition to the feeding and education of individuals or groups with chronic disease risk factors [2]. Primary health care is an effective avenue for the management and prevention of chronic diseases [3]. However, primary health care professionals, including dietitians, report challenges in providing effective care because of time and financial constraints under profession-specific requirements [4-7]. Australian dietitians report restricting the consultation frequency and length because of insufficient funding under the Australian Medicare scheme [4,8]. Such “abbreviated care” is of concern for patient health outcomes [4,8]. Thus, tools to increase the efficiency and effectiveness of practice are important for dietitians working in this sector.

Information technology is recognized as supporting improvements in the health care quality, safety, efficiency, and cost efficiency [7,9,10]. Moreover, information technology facilitates the increased productivity in health care by reducing the time required to complete tasks [11-13]. In the context of the dietetic practice, dietitians can use information technology, such as electronic health records, in consultations for automatic calculations (ie, anthropometry and dietary intake), electronic database storage of patient information, and prepopulated prompting within chart entries (ie, diagnosis prompts based on assessment) [11]. Electronic health records are longitudinal records of patient health information produced from encounters in any health care delivery setting [14]; these can support efficiency, better collaboration between health professions, and provide potential to measure digital health outcomes and inform research to promote better patient outcomes [15,16]. In addition, mobile technology has been used to support dietetic care outside of consultations, such as for motivational reminders and feedback for patients [17-19] and conduct remote dietary tracking [20]. Strategies that support dietitians to utilize information technology are, therefore, likely to have a positive impact on the dietetic practice.

Despite the potential contribution to effective practice, there is limited understanding about the way dietitians use information technology in practice. The opportunity for practice improvements using information technology is particularly relevant to the primary care sector in which >25% of Australian dietitians’ work [21]. Although a body of literature exists investigating dietitians’ use of a mobile app [22,23], a better understanding of the broader use of information technology is needed. Moreover, greater use of information technology, such as electronic health records, by primary care dietitians may facilitate national-scale reporting of patient outcomes to inform policy making and dietetic advocacy. Therefore, this study aims to investigate the perceptions and experiences of Australian primary care dietitians about using information technology in practice.

Methods

Research Overview

This study was underpinned by a descriptive exploratory approach [24] and utilized an inductive qualitative design to capture the perceptions and experiences of Australian primary care dietitians about using information technology in practice [25]. Potential participants were any dietitian working in Australia who had conducted consultations with patients in the primary care setting.

Recruitment

We used purposive and snowball sampling to recruit participants. The recruitment was conducted through direct emails to the members of various dietetic groups, including Dietitian Connection and Hunter Private Practice Dietetics Group, and through posts on dietetic Facebook groups. Participants were asked to forward recruitment materials to other potential participants within their own networks. In addition, participant characteristics (including gender, location, and years of experience) were monitored throughout the recruitment to ensure a broad range of participants, for example, both genders were included.

Data Collection

Data were collected as part of a larger study that also explored primary care dietitians’ perceptions about the effectiveness and efficiency in the workplace. Data collection involved individual semistructured telephonic interviews of 18 and 65 (average, 48) minutes. Data collection was conducted in March 2017 and April 2017. Interviews were conducted using a semistructured interview protocol with open-ended questions that aimed to direct discussion around primary care dietitians’ perceptions and experiences about using information technology in practice. In addition, the interview protocol was developed following a literature review and discussion with a primary care dietitian about the appropriateness of questions. Questions were initially piloted with 2 primary care dietitians within the research team’s network and used to gather feedback on the flow of conversation and provide a model for subsequent interviews. In this study, only the second pilot was included because it accurately reflected the subsequent interviews. Telephonic interviews were recorded using a digital dictaphone and transcribed verbatim. All transcripts were emailed to participants to confirm the accuracy.

Data Analysis

Data analysis was conducted alongside data collection using a thematic analysis approach. Thematic analysis consisted of reading each transcript thoroughly to identify, analyze, and report patterns in the data. Although the first 3 transcripts were analyzed in triplicate, the remainder were analyzed in duplicate. Results from analyses were then compared and combined to ensure that codes accurately reflected participants’ responses. Once the codes were confirmed, themes and subthemes were organized and assigned relevant quotes from the transcripts. In addition, triangulation was conducted with all team members whereby themes and subthemes were reviewed during regular meetings to ensure they reflected the data for transferability and credibility.

Results

Participants' Characteristics

In this study, 20 primary care dietitians (n=17 females) participated in a semistructured interview. Data saturation was reached after 17 interviews because no new codes emerged from the further analysis. Participants were located across 3 states of Australia (New South Wales, Queensland, and Victoria), had a diverse range of experience in primary health care (8 months to 23 years), and most of them worked on a part-time basis. Table 1 summarizes the participants' demographic characteristics.

From the interviews, 4 themes and 10 subthemes emerged, as displayed in Table 2; these were as follows: improving the efficiency of practice tasks, experiencing barriers to using information technology in practice, enhancing outcomes through education and monitoring, and sharing information with others.

In addition, several types of technologies were discussed during interviews. The supplementary material describes their names and purpose.

Theme 1: Information Technology Improving the Efficiency of Practice Tasks

The first theme identified several ways in which information technology could enhance the efficiency of the dietetic practice. Using information technology to aid administrative tasks was viewed as time saving, particularly for writing reports to other health care professionals. As one participant noted:

> *Electronic messaging...to GPs, it's super quick. I've just got standard, self-populating letters that puts in all the basic information and then I can expand on that.* [P16, 10 years' experience]

The time saved using information technology allowed dietitians to further focus on clients' needs:

> *Using electronic notes...my attention is now very much on the client.* [P10, 2 years' experience]

Table 1. The demographic characteristics of the primary care dietitians interviewed in this study.

Interview number	Interview length (min)	Gender	Years qualified as a dietitian	Years working in primary health care	Contracted hours in primary health care	Location of employment	Percentage chronic disease management referrals of total business
1	50	Female	6	1.5	Part time	NSW[a]	0
2	18	Female	4.5	4	Full-time	NSW	0
3	57	Female	6	6	Part time	NSW	90
4	52	Male	15	10	Full time	NSW	0
5	55	Female	8	8	Part time	QLD[b]	90
6	46	Female	2 or 3	0.7	Part time	VIC[c]	50
7	46	Female	5	4	Full time	VIC	85
8	45	Female	30	6	Part time	QLD	90
9	63	Female	33	15	Part time	QLD	95
10	72	Female	3	2	Part time	NSW	40
11	39	Female	3	2.5	Part time	NSW	90
12	56	Female	5	5	Part time	NSW	80
13	44	Female	15	13	Part time	QLD	<10
14	47	Female	9	9	Part time	NSW	20
15	51	Female	3	3	Full time	QLD	90
16	38	Female	25	10	Part time	QLD	90
17	37	Male	14	13	Part time	QLD	90
18	65	Male	7	7	Full time	QLD	95
19	34	Female	28	23	Part time	VIC	60
20	26	Female	29	11	Full time	VIC	20

[a] NSW: New South Wales.

[b] QLD: Queensland.

[c] VIC: Victoria.

Table 2. Themes and subthemes emerging from interviews.

Theme label	Theme description	Subthemes
Improving the efficiency of practice tasks	Information technology can facilitate greater efficiency in dietetic practice	Information technology aids administrative tasks; information technology aids consultation tasks
Experiencing barriers to using information technology in practice	There are numerous barriers to using information technology in the dietetic practice	Implementing new information technology is a low priority; information technology impairs communication with patients; information technology is considered expensive and unreliable
Enhancing outcomes through education and monitoring	Information technology improves patient outcomes by supporting education and tracking progress	Information technology is a valuable tool for patient education; digital tracking makes patients accountable and helps patients achieve goals; information technology increases access to nutrition information
Sharing information with others	Information technology enables sharing of information with other health professionals and with patients outside of consultations	Information technology can increase communication between health professionals; patient outcomes can be shared through digital databases

In addition, information technology was regarded as useful for improving the efficiency of consultation tasks, such as analyzing dietary data and scoring questionnaires concurrently as patients completed them. One dietitian discussed using a program called “Nutritics” to save time:

> *Rather than writing out a food diary for some of my patients, I am actually typing directly into this software system as they tell me what they eat and it gives me a nutritional breakdown of that food.* [P5, 8 years’ experience]

Moreover, participants discussed using information technology for collecting information before consultations, such as through their business website, iPads in office waiting rooms, and food tracking apps. Some participants also used information technology to initiate patient education prior to consultations, reporting that this increased the efficiency of the subsequent consultation; for example,

> *Clients can commence their nutrition education prior to their appointment...this saves time during the consultation. During the consultation, the dietitian can concentrate on consolidation of the education, goal setting and rapport building.* [P9, 15 years’ experience]

Theme 2: Experiencing Barriers to Using Information Technology in Practice

The second theme acknowledged the common barriers experienced by dietitians to using information technology in practice. Barriers that discouraged participants from using information technology included a lack of time to implement new technology, lack of knowledge on available technologies, and a resistance to change. In the quotes below, participants divulge why these barriers discouraged the use of information technology.

> *I have registered for it [health kit] and I’ve looked at it...but it’s just one of those things where changing my processes...getting up to speed on it and having to watch tutorials, etc.* [P3, 6 years’ experience]
>
> *I haven’t really had exposure to e-health records and I don’t know how I would go about setting that up...So, it hasn’t happened...It would be me taking the initiative.* [P14, 9 years’ experience]

Thus, using information technology was a low priority for some participants:

> *It's down the list of priorities. So, it hasn’t happened* [P14, 9 years’ experience]

Others reported that using information technology during the consultation impaired communication with patients. Many dietitians reported having limited control over computer positioning causing them to face away from patients, which in turn led to poor body language, lack of eye contact, and reduced voice audibility. The impact of the information technology use on the voice audibility is articulated by one participant:

> *[The] elderly have hearing problems so we can’t face away from them and talk.* [P5, 8 years’ experience]

In addition, the cost was reported to be a barrier to using information technology in practice:

> *I think there are better systems that we could use but we can’t afford them* [P5, 8 years’ experience]

Moreover, the reliability of devices was reported to limit the use of information technology during consultations,

> *I use my laptop when I go to home visits...it runs out of battery eventually.* [P15, 3 years’ experience]

Furthermore, the internet connection was identified as being unreliable and “*the one that lets you down every single time* ” (P13, 13 years’ experience).

Theme 3: Information Technology Enhancing Outcomes Through Education and Monitoring

The third theme highlighted the potential for improved service delivery and patient outcomes using information technology. Participants identified opportunities for enhanced education using information technology, such as a mobile phone app for celiac disease and fermentable oligo-, di-, mono-saccharides and polyols (FODMAP) diets. These were regarded as important for patients and dietitians to feel confident in identifying foods that are appropriate for consumption.

> *[I can’t] remember every single food, and which one is high, low and medium*

> *in FODMAPs. But [the app gave] the confidence, to show them how to do that.* [P10, 2 years' experience]

The internet was identified as another valuable tool for patient education because it allowed dietitians to promptly access information during consultations and educate patients on finding healthy recipes.

> *If a patient is sitting across from me and I need to look something up then I can just look that up straight away. So, I can have the information on the spot.* [P6, 8 months experience]

Information technology was regarded as a useful means to help patients track their dietary behaviors and progress between consultations. It was identified that the tracking dietary intake helps to empower patients and facilitate better outcomes:

> *Clients get better outcomes when they are empowered by monitoring their own progress. It is vital that they get access to their medical progress information via this technology.* [P9, 15 years' experience]

However, some participants expressed that recording behaviors on apps can be "*burdensome [to patients] because you're having to record so much* " (P3, 6 years' experience). To overcome this challenge, some participants used photos to track food intake, which they considered to be more reliable than self-reported intake:

> *I ask them to take photos of their food if I can't quite judge their portion size* [P15, 3 years' experience]

Participants felt that mobile phone apps improved the accessibility and ease of nutritional information because they are portable. One participant discussed these benefits in terms of the FODMAP app:

> *Information[s] right there when they're shopping...they just download it on their phone [FODMAP app].* [P6, 8 months experience]

However, many participants felt that older patients were less able to use the information technology to access information:

> *The main problem with technology is actually the age group of most of my clients I see a lot of people in their 60s and 70s.* [P8, 6 years' experience]

Theme 4: Information Technology for Sharing Information With Others

The fourth theme acknowledged the potential for greater information sharing using information technology. Most participants used patient management systems to acquire information about patients and communicate with other health care professionals. One participant articulated the use of information technology in enhancing communication:

> *It provides an avenue to communicate with the whole team, so anyone working at the medical centre can see it* [P11, 2.5 years' experience]

Improvements in communication and information sharing were thought to ultimately enhance the continuity of care for patients:

> *The advantages are it means that the patient doesn't feel that they've got to [repeat themselves]...they feel some kind of continuity* [P10, 2 years' experience]

Despite these benefits to patient management systems, many participants expressed a desire for systems that are better suited to the specific information that dietitians collect during consultations:

> *For SGAs [Subjective Global Assessments] we have to do it on paper and then send it to head office and they scan it in and upload it* [P13, 13 years' experience]

Information technology enabled participants to interact with patients outside of the formal consultation. Although some participants actively encouraged patients to email or short message service text message for additional support, others used social media for supplementary nutritional support and sharing recipe ideas or emerging dietary evidence.

> *[I] encourage clients to email and text me if they have questions...The closed Facebook groups [allow me to] give clients, gentle reminders and keep them up to speed with recipe ideas.* [P12, 5 years' experience]

Participants acknowledged the potential for information technology to facilitate the collation of data and evaluate the effectiveness of dietetic services:

> *It would be beneficial in being able to determine the effectiveness of private practice dietitians* [P1, 1.5 years' experience]

Moreover, participants expressed willingness to share patient outcomes utilizing a digital database but were concerned about the confidentiality of information and the ability of a digital database to integrate with current practice software:

> *It could be doubling up...putting my information into two databases...I don't have time to do two things* [P16, 10 years' experience]

Discussion

This study was the first to use a qualitative methodology to explore Australian primary care dietitians' perceptions about using information technology during practice. Dietitians in this study viewed information technology as beneficial to their practice and important for enhancing patient outcomes. Understanding these perceptions provides a direct insight into whether information technology is feasible to use in the dietetic practice and opportunities for greater integration of technology in the future.

Participants in this study regarded information technology as important for enhancing the efficiency of administrative and consultation tasks within the dietetic practice; these included administrative technology, mobile phone apps for tracking dietary intake, patient management systems, and Nutritics (a data analysis system). Technological devices have previously been shown to reduce dietitian workloads [11,13,26]; for example, using an iPad compared with paper-based forms may reduce consultation time by 5 minutes [27] and using electronic systems may reduce consultation time by 13 minutes [11].

Furthermore, an estimated 40% of preconsultation assessments can be completed remotely using information technology [27]. There are clear benefits of information technology use for improving the consultation efficiency.

Despite benefits to efficiency, better integration of dietary assessments with current health care electronic systems is needed. Interviews from this study indicate that the utilization of preconsultation data collection is not optimal with a digital assessment of the dietary intake and patient information being obtained separately. However, one participant utilized an independent app, NERO [28], which integrated the collection of dietary and medical assessments. Nevertheless, a literature review does not identify any similar app widely available. In addition, other studies acknowledge the needs of dietitians with regard to the technology design, such as linking app data with electronic health records for better work efficiency [29,30]. However, most studies investigate the app quality and patient usability [22,31] as opposed to dietitian preferences to support practice tasks. In 2013, a survey of Australian employed dietitians reported that 77% had no experience with nutrition-related information technology systems [32]. It has previously been identified that education, training, and advocacy of technology needs to be provided by dietetic associations to encourage better use or development of information technology within the profession [29,33]. Clearly, there is a need for better promotion or further development of digital assessments in health care to accurately reflect dietitians' needs and, therefore, further enhance the efficiency.

Dietitians in this study felt that electronic patient management systems support better communication and sharing of patient information. Similar benefits of patient management systems have been previously highlighted through surveys of dietitians [5]. However, participants in this research acknowledged that these systems were not well integrated with dietitian needs. Currently, there are over 20 commercial systems available to primary care dietitians that do not align with the nutrition care process [34]. Considering that most participants reported using patient management systems in their practice, further work is needed to streamline the dietetic documentation.

Participants regarded information technology as useful for education and monitoring progress. Smartphone apps have been shown to promote self-monitoring by reducing the burden of recording [20]. In addition, the availability of barcode scanning, nutrient databases, and image recording contributes to the reduced burden [35,36]. Studies have reported that the inclusion of features in apps that are user friendly and reduce time burden is more effective [37]. A review of 23 studies showed that smartphone apps are beneficial in targeting behavioral change (17 studies) and increasing the retention rate for interventions (19 studies) [37]. In light of this, dietitians worldwide use information technology for patient self-monitoring [22,23], and dietitians in United Kingdom are recommended to use information technology to support their practice [38]. Overall, the use of information technology in the dietetic practice appears to be well accepted and encouraged internationally.

In 2014, the Academy of Nutrition and Dietetics Health Informatics Infrastructure was launched as a tool that enables documentation and standardized data collection for outcomes research [39]. The Academy of Nutrition and Dietetics Health Informatics Infrastructure system is based on the nutrition care process and international dietetic terminology [39]. This research shows that using information technology for tracking standardized data on nutrition-related outcomes could be beneficial for promoting evidence-based improvements in nutrition care. Similar benefits have been identified among other health professions, such as for identifying survival rates and characteristics among never smokers for patients with lung cancer [40]. Systems that can systematically document outcomes may be beneficial to dietitians by pooling data to evaluate the impact of structures and processes on patient outcomes. Most participants expressed interest or willingness to enter data using similar technologies. Likewise, focus group discussions assessing dietitians' perceptions of a prototype Web-based electronic recording system indicated a possibility that it was easy to use and would be useful in patient management systems [41]. Alternate views are represented in surveys, showing that Australian dietitians lack readiness for eHealth [30]. However, survey results between 2013 and 2016 showed an improvement for eHealth readiness [30], which suggests further acceptance from the dietetic profession regarding the eHealth implementation. Therefore, this study supports trailing standardized data collection using information technology within the broader dietetic population.

There are some notable limitations to this study. First, interviews were conducted over the phone, which limited the ability to monitor body language and other interpersonal cues. However, utilizing telephonic interviews to collect data extended the geographical access of participants and allowed a broad range of individuals to be represented. Although Western Australia, Tasmania, and South Australia were not included in the study sample, this reflects the higher proportion of employed dietitians within the other states of Australia [42]. Second, the interviews were conducted by 3 interviewees, which may have reduced the consistency among interviews. However, all interviewees were concurrently trained in the interview protocol and were present at each interview, enabling prompting, which promoted consistency among participants.

Information technology appears to be an important and acceptable component of the dietetic practice in primary care. This study adds insight into the variety of ways in which information technology can benefit dietitians, while also identifying several barriers to overcome for its optimal use. The potential benefits of information technology include better communication and information sharing among health care professionals to support patient outcomes, improved access to information to support education and interventions, digital dietary tracking to encourage patient accountability and better outcomes, better communication and accessibility of dietetic services (ie, Facebook groups); and potential to support improvements in nutrition care through standardized databases and outcomes research.

The opportunities for improved patient outcomes through information technology use warrant strategies to support dietitians overcome barriers to integrating technology into practice. These barriers include technology detracting from the patient-dietitian communication, technology not being appropriate for certain patient groups, or tasks and technology not integrating well into dietetics. It is important that employers support dietitians by providing training on effectively using technology in consultations or setting up consultation rooms to support better body language. Technology is not suitable for some patient populations; therefore, dietitians should also be mindful of the alternative methods of educating and collecting data from these individuals. To accurately reflect dietitians' needs, relevant nutrition resources, such as subjective global assessments and the nutrition care process, need to be included into patient management systems. Furthermore, linking app data with electronic health records or patient management systems is needed. The successful development and adoption of such information technologies will require support from dietitians and national bodies, such as the Dietitian Association of Australia and the Australian Government. Further development of information technology is needed to better support dietitians in the workplace.

Acknowledgments

This project was conducted as part of a Bachelor of Nutrition & Dietetics Honours Program. Griffith University provided funds to cover project expenses such as the cost of vouchers.

Authors' Contributions

AJ participated in the acquisition, analysis, interpretation of data, and drafting of work. RC, KS, and LTW participated in the acquisition and analysis of data. LJM, MER, LB, and LTW substantially contributed to the research concept and revision of work.

Conflicts of Interest

None declared.

Multimedia Appendix 1

A descriptive list of the technologies dietitians use to aid their practice.

[PDF File (Adobe PDF File), 83KB - jmir_v20i10e265_app1.pdf]

References

1. World Health Organization. Integrated chronic disease prevention and control. 2017. URL: http://www.who.int/chp/about/integrated_cd/en/ [accessed 2018-09-27]
2. Dietitians Association of Australia. What Dietitians Do. 2017. URL: https://daa.asn.au/what-dietitans-do/dietitian-or-nutritionist/ [accessed 2017-12-01] [WebCite Cache ID 6vNabG9JB]
3. Australian Institute of Health and Welfare. Primary health care in Australia Canberra Australian Institute of Health and Welfare. 2014. URL: https://www.aihw.gov.au/reports-statistics [accessed 2017-12-01] [WebCite Cache ID 6vNaqnPQZ]
4. Foster MM, Cornwell PL, Fleming JM, Mitchell GK, Tweedy SM, Hart AL, et al. Better than nothing? Restrictions and realities of enhanced primary care for allied health practitioners. Aust. J. Prim. Health 2009;15(4):326-334. [doi: 10.1071/PY08065]
5. Cant RP. Patterns of delivery of dietetic care in private practice for patients referred under Medicare Chronic Disease Management: results of a national survey. Aust Health Rev 2010 May;34(2):197-203. [doi: 10.1071/AH08724] [Medline: 20497733]
6. Foster M, Mitchell G, Haines T, Tweedy S, Cornwell P, Fleming J. Does enhanced primary care enhance primary care? Policy-induced dilemmas for allied health professionals. Med J Aust 2008 Jan 7;188(1):29-32. [Medline: 18205560]
7. Carey M, Noble N, Mansfield E, Waller A, Henskens F, Sanson-Fisher R. The Role of eHealth in Optimizing Preventive Care in the Primary Care Setting. J Med Internet Res 2015 May;17(5):e126 [FREE Full text] [doi: 10.2196/jmir.3817] [Medline: 26001983]
8. Cant RP. Public health nutrition: the accord of dietitian providers in managing medicare chronic care outpatients in Australia. Int J Environ Res Public Health 2010 Dec;7(4):1841-1854 [FREE Full text] [doi: 10.3390/ijerph7041841] [Medline: 20617063]
9. Buntin MB, Burke MF, Hoaglin MC, Blumenthal D. The benefits of health information technology: a review of the recent literature shows predominantly positive results. Health Aff (Millwood) 2011 Mar;30(3):464-471 [FREE Full text] [doi: 10.1377/hlthaff.2011.0178] [Medline: 21383365]
10. DesRoches CM, Campbell EG, Rao SR, Donelan K, Ferris TG, Jha A, et al. Electronic health records in ambulatory care--a national survey of physicians. N Engl J Med 2008 Jul 03;359(1):50-60. [doi: 10.1056/NEJMsa0802005] [Medline: 18565855]
11. Rossi M, Campbell KL, Ferguson M. Implementation of the Nutrition Care Process and International Dietetics and Nutrition Terminology in a single-center hemodialysis unit: comparing paper vs electronic records. J Acad Nutr Diet 2014 Jan;114(1):124-130. [doi: 10.1016/j.jand.2013.07.033] [Medline: 24161368]

12. Miskulin DC, Weiner DE, Tighiouart H, Ladik V, Servilla K, Zager PG, Medical Directors of Dialysis Clinic Inc. Computerized decision support for EPO dosing in hemodialysis patients. Am J Kidney Dis 2009 Dec;54(6):1081-1088 [FREE Full text] [doi: 10.1053/j.ajkd.2009.07.010] [Medline: 19781831]
13. Mickan S, Tilson JK, Atherton H, Roberts NW, Heneghan C. Evidence of effectiveness of health care professionals using handheld computers: a scoping review of systematic reviews. J Med Internet Res 2013;15(10):e212 [FREE Full text] [doi: 10.2196/jmir.2530] [Medline: 24165786]
14. Health Information and Management Systems Society. Electronic Health Records. 2017. URL: https://www.himss.org/library/ehr [accessed 2018-04-19] [WebCite Cache ID 6yjgFIhte]
15. Houser H, Johnson A. Perceptions Regarding Electronic Health Record Implementation among Health Information Management Professionals in Alabama: A Statewide Survey and Analysis. Perspectives in Health Information Management 2008;5(6):1-15. [Medline: 18504505]
16. National National E-Health Transition Authority. Getting Started With Digital Health: Benefits. 2017. URL: https://www.digitalhealth.gov.au/get-started-with-digital-health/benefits [accessed 2018-04-17] [WebCite Cache ID 6yjhKrZtP]
17. Burke LE, Conroy MB, Sereika SM, Elci OU, Styn MA, Acharya SD, et al. The effect of electronic self-monitoring on weight loss and dietary intake: a randomized behavioral weight loss trial. Obesity (Silver Spring) 2011 Feb;19(2):338-344 [FREE Full text] [doi: 10.1038/oby.2010.208] [Medline: 20847736]
18. Patrick K, Raab F, Adams MA, Dillon L, Zabinski M, Rock CL, et al. A text message-based intervention for weight loss: randomized controlled trial. J Med Internet Res 2009 Jan;11(1):e1 [FREE Full text] [doi: 10.2196/jmir.1100] [Medline: 19141433]
19. Partridge S, McGeechan K, Hebden L, Balestracci K, Wong A, Denney-Wilson E, et al. Effectiveness of a mHealth Lifestyle Program With Telephone Support (TXT2BFiT) to Prevent Unhealthy Weight Gain in Young Adults: Randomized Controlled Trial. JMIR Mhealth Uhealth 2015 Jun 15;3(2):e66 [FREE Full text] [doi: 10.2196/mhealth.4530] [Medline: 26076688]
20. Hutchesson MJ, Rollo ME, Callister R, Collins CE. Self-monitoring of dietary intake by young women: online food records completed on computer or smartphone are as accurate as paper-based food records but more acceptable. J Acad Nutr Diet 2015 Jan;115(1):87-94. [doi: 10.1016/j.jand.2014.07.036] [Medline: 25262244]
21. Ball L, Larsson R, Gerathy R, Hood P, Lowe C. Working profile of Australian private practice Accredited Practising Dietitians. Nutrition & Dietetics 2013 Jan 29:n/a-n/a. [doi: 10.1111/1747-0080.12015]
22. Lieffers JRL, Vance VA, Hanning RM. Use of mobile device applications in Canadian dietetic practice. Can J Diet Pract Res 2014 Mar;75(1):41-47. [doi: 10.3148/75.1.2014.41] [Medline: 24606959]
23. Jospe MR, Fairbairn KA, Green P, Perry TL. Diet app use by sports dietitians: a survey in five countries. JMIR Mhealth Uhealth 2015 Jan 22;3(1):e7 [FREE Full text] [doi: 10.2196/mhealth.3345] [Medline: 25616274]
24. Smith J. Interpretative Phenomenological Analysis: Theory, Method and Research. Thousand Oaks, United States: Sage Publications Ltd; 2009.
25. Liamputtong P. Research methods in health: foundations for evidence-based practice. South Melbourne, Victoria, Australia: Oxford University Press; 2013.
26. Ngo J, Engelen A, Molag M, Roesle J, García-Segovia P, Serra-Majem L. A review of the use of information and communication technologies for dietary assessment. Br J Nutr 2009 Jul;101 Suppl 2:S102-S112. [doi: 10.1017/S0007114509990638] [Medline: 19594959]
27. Howell M, Hood AJ, Jayne DG. Use of a patient completed iPad questionnaire to improve pre-operative assessment. J Clin Monit Comput 2017 Feb;31(1):221-225. [doi: 10.1007/s10877-015-9818-0] [Medline: 26715416]
28. NERO. NERO - nutrition education resources online. 2017. URL: http://www.nero4me.com.au/nero4me.html [accessed 2017-12-01] [WebCite Cache ID 6vNbi5xpH]
29. Chen J, Lieffers J, Bauman A, Hanning R, Allman-Farinelli M. Designing Health Apps to Support Dietetic Professional Practice and Their Patients: Qualitative Results From an International Survey. JMIR Mhealth Uhealth 2017 Mar 31;5(3):e40. [doi: 10.2196/mhealth.6945]
30. Maunder KW, Walton K, Williams P, Ferguson M, Beck E. eHealth readiness of dietitians. J Hum Nutr Diet 2018 Feb 23:573-583. [doi: 10.1111/jhn.12542] [Medline: 29473238]
31. Chen J, Cade JE, Allman-Farinelli M. The Most Popular Smartphone Apps for Weight Loss: A Quality Assessment. JMIR Mhealth Uhealth 2015 Dec 16;3(4):e104 [FREE Full text] [doi: 10.2196/mhealth.4334] [Medline: 26678569]
32. Maunder K, Walton K, Williams P, Ferguson M, Beck E, Ayres E, et al. Uptake of nutrition informatics in Australia compared with the USA. Nutrition & Dietetics 2015 Jun 18;72(3):291-298. [doi: 10.1111/1747-0080.12207]
33. Chen J, Lieffers J, Bauman A, Hanning R, Allman-Farinelli M. The use of smartphone health apps and other mobile health (mHealth) technologies in dietetic practice: a three country study. J Hum Nutr Diet 2017 Jan 24:439-452. [doi: 10.1111/jhn.12446] [Medline: 28116773]
34. Probst Y. Dietitians in the electronic age: Progressing towards e-health. Nutrition & Dietetics 2011;68(3):177-178. [doi: 10.1111/j.1747-0080.2011.01542.x]
35. Warren JI, McLaughlin M, Bardsley J, Eich J, Esche CA, Kropkowski L, et al. The Strengths and Challenges of Implementing EBP in Healthcare Systems. Worldviews Evid Based Nurs 2016 Feb;13(1):15-24. [doi: 10.1111/wvn.12149] [Medline: 26873372]

36. Walton J. Dietary Assessment Methodology for Nutritional Assessment. Topics in Clinical Nutrition 2015;30(1):33-46. [doi: 10.1097/TIN.0000000000000018]
37. Zhao J, Freeman B, Li M. Can Mobile Phone Apps Influence People's Health Behavior Change? An Evidence Review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
38. Department of Health. A Framework for Technology Enhanced Learning. 2011. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/215316/dh_131061.pdf [accessed 2017-12-01] [WebCite Cache ID 6vNbvaalN]
39. Murphy WJ, Steiber AL. A new breed of evidence and the tools to generate it: introducing ANDHII. J Acad Nutr Diet 2015 Jan;115(1):19-22. [doi: 10.1016/j.jand.2014.10.025] [Medline: 25534894]
40. Clément-Duchêne C, Stock S, Xu X, Chang ET, Gomez SL, West DW, et al. Survival among Never-Smokers with Lung Cancer in the Cancer Care Outcomes Research and Surveillance Study. Ann Am Thorac Soc 2016 Jan;13(1):58-66 [FREE Full text] [doi: 10.1513/AnnalsATS.201504-241OC] [Medline: 26730864]
41. O'Sullivan TA. Evaluation of an electronic record prototype incorporating the Nutrition Care Process and International Dietetics and Nutrition Terminology. Nutrition & Dietetics 2013 Jan 29;70(3):188-195. [doi: 10.1111/1747-0080.12012]
42. Health Workforce Australia. Australia’s Health Workforce Series Dietitians in Focus. 2014. URL: http://iaha.com.au/wp-content/uploads/2014/03/HWA_Australias-Health-Workforce-Series_Dietitians-in-focus_vF_LR.pdf [accessed 2017-12-01] [WebCite Cache ID 6vNc0PmYY]

Abbreviations

FODMAP: fermentable oligo-, di-, monosaccharides and polyols

Edited by G Eysenbach; submitted 04.12.17; peer-reviewed by J Chen, R Pankomera; comments to author 17.03.18; revised version received 19.04.18; accepted 24.06.18; published 15.10.18.

Please cite as:
Jones A, Mitchell LJ, O'Connor R, Rollo ME, Slater K, Williams LT, Ball L
Investigating the Perceptions of Primary Care Dietitians on the Potential for Information Technology in the Workplace: Qualitative Study
J Med Internet Res 2018;20(10):e265
URL: https://www.jmir.org/2018/10/e265/
doi:10.2196/jmir.9568
PMID:30322837

Original Paper

Web-Based Information Infrastructure Increases the Interrater Reliability of Medical Coders: Quasi-Experimental Study

Julian Varghese[1], MD, MSc; Sarah Sandmann[1], PhD; Martin Dugas[2], MD, MSc

[1] Institute of Medical Informatics, University of Münster, Münster, Germany

[2] Institute of Medical Informatics, European Research Center for Information Systems, Münster, Germany

Corresponding Author:
Julian Varghese, MD, MSc
Institute of Medical Informatics
University of Münster
Albert-Schweitzer-Campus 1
Gebäude A11
Münster, 48149
Germany
Phone: 49 2518354714
Fax: 49 251 83 52259
Email: julian.varghese@ukmuenster.de

Abstract

Background: Medical coding is essential for standardized communication and integration of clinical data. The Unified Medical Language System by the National Library of Medicine is the largest clinical terminology system for medical coders and Natural Language Processing tools. However, the abundance of ambiguous codes leads to low rates of uniform coding among different coders.

Objective: The objective of our study was to measure uniform coding among different medical experts in terms of interrater reliability and analyze the effect on interrater reliability using an expert- and Web-based code suggestion system.

Methods: We conducted a quasi-experimental study in which 6 medical experts coded 602 medical items from structured quality assurance forms or free-text eligibility criteria of 20 different clinical trials. The medical item content was selected on the basis of mortality-leading diseases according to World Health Organization data. The intervention comprised using a semiautomatic code suggestion tool that is linked to a European information infrastructure providing a large medical text corpus of >300,000 medical form items with expert-assigned semantic codes. Krippendorff alpha (K_{alpha}) with bootstrap analysis was used for the interrater reliability analysis, and coding times were measured before and after the intervention.

Results: The intervention improved interrater reliability in structured quality assurance form items (from K_{alpha}=0.50, 95% CI 0.43-0.57 to K_{alpha}=0.62 95% CI 0.55-0.69) and free-text eligibility criteria (from K_{alpha}=0.19, 95% CI 0.14-0.24 to K_{alpha}=0.43, 95% CI 0.37-0.50) while preserving or slightly reducing the mean coding time per item for all 6 coders. Regardless of the intervention, precoordination and structured items were associated with significantly high interrater reliability, but the proportion of items that were precoordinated significantly increased after intervention (eligibility criteria: OR 4.92, 95% CI 2.78-8.72; quality assurance: OR 1.96, 95% CI 1.19-3.25).

Conclusions: The Web-based code suggestion mechanism improved interrater reliability toward moderate or even substantial intercoder agreement. Precoordination and the use of structured versus free-text data elements are key drivers of higher interrater reliability.

(J Med Internet Res 2018;20(10):e274) doi:10.2196/jmir.9644

KEYWORDS

clinical coding; health information interoperability; Unified Medical Language System; eligibility criteria

Introduction

The rise of electronic documentation in health care and research aims to improve patient data exchange not only for proper payment or reimbursement but also for improved data analysis and patient safety; this produces thousands of terabytes of data annually in the United States and Europe [1]. However, ineffective workflows, heterogeneity, and redundancy of data affect the data quality [2-4] and hamper its reuse, comparison, and analysis across different research institutions. Nonstructured and structured data are affected because data elements might be defined or interpreted differently. Semantic coding of data elements enables the identification of semantically matching elements in different data sources and is a key step toward data integration [5,6] and enables the generation of disease-specific core datasets for efficient data capture [7]. Natural Language Processing (NLP) tools use semantic codes for dictionary look-up algorithms or normalize medical terms in clinical free-text notes and use existing semantic thesaurus relations to infer semantic analyses of text segments [8-12].

All of the mentioned examples [5-12] had used the largest clinical metathesaurus available to code medical concepts, the Unified Medical Language System (UMLS) [13], which currently contains >3 million unique concepts; it includes several biomedical vocabularies, for example, clinical reference terminologies as SNOMED Clinical Terms (SNOMED-CT) or Logical Observation Identifiers Names and Codes and medical classifications such as the ICD-10 or other well-known coding systems such as MeSH or MeDRA. Besides the aforementioned advantages of semantic coding of medical content, there is an issue regarding ambiguity in expert-based assignments of such semantic codes—some of the many concepts are synonymous but are given different concept identifiers; for example, the UMLS concept “antidementia drug” (Code: C1276997) and the concept “antidementia agents” (Code: C1531592) are synonymous from a clinical point of view but are represented by 2 different codes. In addition, there is a low semantic similarity or relatedness between those concepts based on the ontological structure of UMLS source vocabularies [14].

In practice, synonymy and abundance of very similar but different concepts in some clinical subdomains lead to inconsistent coding among different coding experts, which weakens the advantage of semantic coding to improve data integration. NLP or information extraction tools that use UMLS as core terminology are affected as well because their programmed assumption “Different UMLS codes represent semantically different concepts” is flawed. Moreover, the evaluation of different NLP tools is challenging, that is, if 2 NLP tools A and B suggested 2 different UMLS codes for a text segment, it may not be clear whether the output of A or B is more valid.

Low rates of human expert-based interrater reliability have been evaluated for the clinical terminology SNOMED-CT [15,16]. Rothschild et al [17] reported that interrater reliability is moderate at best with UMLS in unstandardized problem lists and suggested that coder training and standardization might improve interrater reliability. Our work provides a significant novelty by performing a systematic interrater reliability analysis of UMLS on several mortality-leading disease domains and conduction of an interventional pre-post study. The intervention included a coder training for using a Web-based semiautomatic code suggestion tool that utilizes a large metadata repository as an expert coding knowledgebase to improve uniform coding among different raters. Following key questions formed the rationale of this study:

1. What is the effect of the intervention on interrater reliability?
2. How does interrater reliability differ when coding structured data elements versus free-text?
3. How does interrater reliability relate to precoordinated versus postcoordinated concepts?

Question 1 seeks to improve interrater reliability by systemizing the way a coder uses a large metadata registry to reuse common precoded medical concepts. This way, a coder would be suggested a preferred UMLS code based on the coding frequency of other expert coders who have already coded the same or similar portion of text.

As for question 2, it is well known that structured data elements are more suitable for data exchange across different information systems than free-text in clinical reports or eligibility criteria of clinical trials [18]. However, free text is still existing and necessary in medical documentation. Both structured data elements and free-text elements could be semantically annotated to foster semantic interoperability. Therefore, our study measured interrater reliability in structured routine documentation forms versus free-text eligibility criteria to examine interrater reliability differences depending on both types of documentation.

Regarding question 3, to code medical concepts of medical data elements or free text, one distinguishes 2 basic semantic coding methods, called precoordination and postcoordination [19]. A medical concept is precoordinated if its semantics are represented by one semantic code; for example, the term “Patient has diabetes mellitus type 2” contains the medical concept “diabetes mellitus type 2” and can be coded by a single UMLS concept code “C0011860-Diabetes Mellitus, Non-Insulin-Dependent.” A medical concept is postcoordinated if it is coded by multiple codes to express more complex semantics; for example, the term “Patient has an allergy to Amoxicillin” contains the medical concept “Allergy to Amoxicillin” and could be coded by the following 2 interrelated UMLS codes: “C0020517-Hypersensitivity” and “C0002645- Amoxicillin.” Another type of coding that can be considered as a special type of postcoordination is the coding of multiple separated medical concepts in one medical term; for example, “Patient has an allergy to Amoxicillin or has diabetes mellitus type 2” contains 2 preceding medical concepts and could represent a free-text inclusion criterion in clinical trials.

The assessment of semantic coding correctness among different coders was not the scope of this analysis and would be pointless because coders recruited for the studies were considered as medical terminology experts. Therefore, the key challenge for accurate UMLS coding is to achieve high interrater reliability among different coders rather than finding a semantically correct coding.

To address these 3 key questions, a quasi-experimental study was conducted with medical experts as study subjects to report on the effects on the coding behavior with and without a Web-based coding suggestion tool.

Methods

Information Infrastructure and Recruitment

The Medical Data Models (MDM) portal is a Web-based large, open-access, metadata registry and European information infrastructure [20] funded by the German Research Federation. More than 15,000 medical forms with >300,000 form items are available; all of them are UMLS coded by medical experts. Each expert undergoes training on how to use an expert-based code suggestion mechanism [5] within the MDM portal; by this, each new coder can choose from previously coded concepts if similar item text patterns exist. To increase the throughput of coded items for the MDM project, 6 final-year medical students from the medical faculty in Münster, Germany, were recruited. None of the students had any experience in UMLS coding. Interrater reliability was assessed before and after the training to address the aforementioned study key questions.

Study Setting and Material

The study is a pre-post analysis that was conducted from February 15, 2017 to March 12, 2017, at the Institute of Medical Informatics, University of Münster (Münster, Germany). Figure 1 illustrates the flowchart of the study design; each coder coded a single form per day and days are regarded as consecutive working days.

We randomly selected 10 eligibility criteria forms of different clinical trials (conducted between 2000 and 2016) from ClinicalTrials.gov. Based on the manual review, each form was only selected if its study was related to a medical condition that was a leading cause of death based on the World Health Organization 2015 Global Health estimate data [21]. If a form did not adhere to this, it was discarded, and the next form was considered until a set of 10 forms was selected. Then, 8 quality assurance forms were analogously collected for the preinterventional phase to provide a dataset of structured documentation forms with a similar number of items. The quality assurance forms originated from the Institute for Applied Quality Improvement and Research in Health Care [22] in Germany and Austria and are implemented by law in all hospitals of Germany that provide therapeutic procedures, which are under governmental quality assurance [23,24]. These forms contain a series of structured routine documentation items, including quality indicators before, during, and after health care procedures.

Figure 1. The study workflow. EC: eligibility criteria; QA: quality assurance.

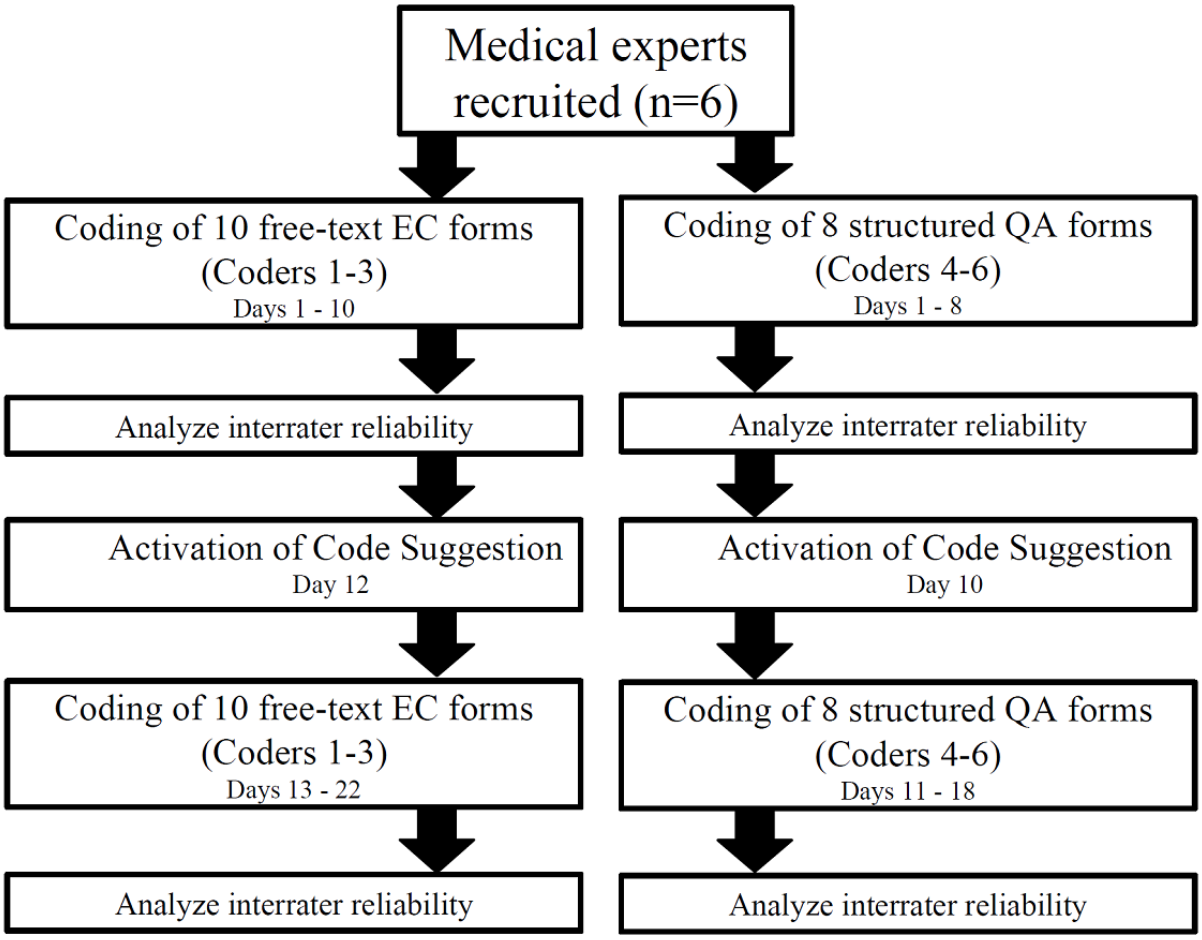
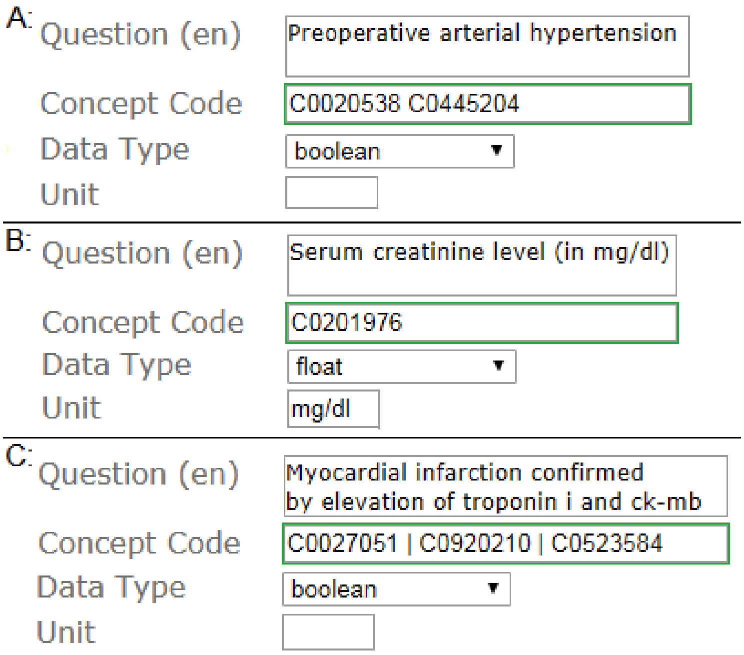


Table 1. The disease category coverage in eligibility criteria forms and quality assurance forms.

Disease category	Documentation models to code (number form models)			
	Eligibility criteria pre-intervention	Eligibility criteria post-intervention	Quality assurance pre-intervention	Quality assurance post-intervention
Cardiovascular (including myocardial infarction and stroke)	4	4	2	2
Respirational diseases	3	3	1	1
Diabetes mellitus and pancreatic diseases	1	1	1	1
Renal diseases	0	0	1	1
Liver diseases	0	0	1	1
Breast cancer	0	0	1	1
HIV/AIDS	1	1	0	0
Traumatic or orthopedic diseases	1	1	1	1

Figure 2. Item coding view in ODMEdit.



A:

Question (en) Preoperative arterial hypertension

Concept Code C0020538 C0445204

Data Type boolean

Unit

B:

Question (en) Serum creatinine level (in mg/dl)

Concept Code C0201976

Data Type float

Unit mg/dl

C:

Question (en) Myocardial infarction confirmed by elevation of troponin i and ck-mb

Concept Code C0027051 | C0920210 | C0523584

Data Type boolean

Unit

For every form in the preinterventional phase, a different disease-matching form was collected for the postinterventional phase. Therefore, 36 forms (2×10 eligibility criteria forms+2×8 quality assurance forms) were included in total for this study. Table 1 summarizes the medical condition categories.

All selected forms were transformed to the Operational Data Model (ODM), an international standard by the Clinical Data Interchange Standards Consortium [25] for the representation and exchange of clinical trial data and metadata. In addition, semantic coding was added using ODMEdit [5]—a Web-based app, which is accessible from the MDM portal and features the semiautomatic code suggestion based on coded items in the MDM database. Multimedia Appendix 1 provides the full set of all study forms (eligibility criteria + quality assurance forms) as Clinical Data Interchange Standards Consortium ODM files. Figure 2 shows screenshots of 3 exemplary items in the item coding view that every coder used. All 3 items covered the 3 aforementioned types of concept coding (precoordination, postcoordination, and multiple concept coding). Figure 2 shows the quality assurance item that needed postcoordination (2 codes to represent one complex concept: preoperative arterial hypertension) and the quality assurance item that was coded with one precoordinated concept with one Unified Medical Language System code: C0201976—Serum creatinine level. Measurement units, data type, or permissible values were not the scope of semantic coding. It also shows the eligibility criteria form item expressing an eligibility criterion of a clinical trial that was coded with 3 codes to express 3 different medical concepts (myocardial infarction as a diagnosis, troponin I, and ck-mb as laboratory markers).

All 6 recruits agreed to participate and were then randomly assigned to code eligibility criteria forms of clinical trials or quality assurance forms. In the preinterventional phase, all coders received a basic 15-minute introduction to use the standard UMLS metathesaurus browser (version 2015AB) to understand the concepts of pre- and postcoordination and use ODMEdit for semantic annotation of form items. In addition, they were instructed to UMLS code-relevant medical concepts of their given form items. Relevant medical concepts were defined to be concepts that the coder deemed significant to capture the semantics of the given form item. Of note, UMLS coding was restricted to the concept domain of possible medical data elements; it does not concern the coding of the value domain. For instance, in the term “Creatinine value of >7 mg/dL,” “Creatinine” is the medical concept of the concept domain; “>7 mg/dL” represents the corresponding value and measurement unit and, therefore, is not considered for the coding procedure. The use of the integrated code suggestion function—which is the key part of the intervention—was prohibited in the preinterventional phase. All recruits performed UMLS coding at their homes using a broadband internet connection to access the MDM portal.

Study Intervention

The intervention of this study consisted of a 60-minute coder training to teach the use of ODMEdit’s code suggestion function and basic coding principles [18]. Each training session consisted of one participant and the same supervisor with extensive experience in UMLS coding in structured data elements [7] and free-text eligibility criteria [18]. The details of semiautomatic ODMEdit’s code suggestion tool are described in our previous work [5]. Multimedia Appendix 1 provides the full teaching material that was provided to the coders.

Each coder was free to follow the coding suggestion (or not). None of the study forms were part of the training, and all participants were prohibited from sharing or discussing their coded forms during the study period. None of the 36 study forms existed in the MDM database before the start of the study. If a coder did not find a suitable code through the suggestion function, the code search could be extended with the standard UMLS metathesaurus browser. Notably, no time restriction was applied in the pre- and postinterventional phase. If coders ended up with no suitable code for a form item, it was up to them to skip the item and move on to the next one.

Interrater Reliability Measure and Coding Time

UMLS codes represent nominal data. The interrater reliability statistics as the simple calculation of percentages of the observed agreement are associated with biases and should be corrected for the agreement expected by chance [26]. Further measures as Cohen kappa [27] are restricted to the use of 2 raters or other limitations [28-31], whereas Krippendorff alpha (K_{alpha}) [32] with bootstrap CIs is reported as recommended especially in cases of missing data—compared with Fleiss K—with >2 raters and with a large amount of different rating categories [31]. Landis and Koch [33] proposed the following interpretation regarding K_{alpha} value ranges: <0, poor agreement; 0.00-0.20, slight agreement; 0.21-0.40, fair agreement; 0.41-0.60, moderate agreement; 0.61-0.80, substantial agreement; and 0.81-1.00, almost perfect agreement.

K_{alpha} was calculated among all 3 raters in all 4 subgroups as follows: preinterventional eligibility criteria forms; postinterventional eligibility criteria forms; preinterventional quality assurance forms; and postinterventional quality assurance forms. The coding of a form item between 2 coders is only considered to be matching if the set of UMLS codes were the same. The calculation was performed with the R-package by Zapf et al [31]; this package also includes bootstrap analysis to determine CIs using 10,000 bootstraps.

Each coder measured his or her coding time via stopwatch by starting the time before the first item and stopping after the last item of a form. Thus, data on the mean coding time per item for each form are available but without item-based time variances. The median coding time per item was then determined for each user as the median mean coding time per item of all forms before and after the intervention.

Of note, this study did not intend to analyze the intervention as a cause of coding time differences. For this purpose, a different study design would have been appropriate. Instead, this study was designed to primarily analyze the effects on interrater reliability. However, time measurements were still taken to audit any adverse coding time expenditures associated with the intervention. As each coder could potentially remember same or similar items from previous forms, a learning effect could bias the time measurements for consecutive forms. To account for this issue, time measurements will be presented for each coder coupled with a learning graph to illustrate the number of new medical concepts for each coding day of the study.

Results

Effect of the Intervention on the Interrater Reliability and Coding Time

K_{alpha} increased for both documentation types (structured quality assurance and free-text eligibility criteria). A significant difference with respect to 95% CIs existed within the eligibility criteria study group (0.43, 95% CI 0.37-0.50 vs 0.19, 95% CI 0.14-0.24). The median word counts per item were comparable for pre- and postinterventional form item sets based on the interquartile ranges (see Table 2 for details).

Based on time measurements for each coder, median coding time per item was decreased in all 6 raters but with overlapping interquartile ranges in the quality assurance subgroup, as shown in Figure 3; it includes the median with interquartile ranges is calculated on the basis of mean coding time per items of all forms before and after the intervention; coders 1-3, free-text eligibility criteria study group and coders 4-6, structured quality assurance forms. Figures 4 and 5 illustrate the coding time during the course of the full study coupled with the aforementioned graph of newly coded concepts on each day for eligibility criteria and quality assurance forms, respectively. On each day, identical form items were coded among different coders. The mean intercoder time difference averages absolute time differences between all 3 coders on each day, as shown in Figure 4; it also includes the interception and slope of each coder-related learning graph is calculated on the basis of linear regression. Within the preinterventional eligibility criteria forms, the median number of new medical concepts among all raters was 20 per day (interquartile range, IQR 16.25-24). After the intervention, it reduced to 13 (IQR 9.25-17).

Table 2. The effect of the intervention on the interrater reliability.

Coded models	Preintervention			Postintervention		
	Number of items	MWC[a] (IQR[b])	K_{alpha}[c] (95% CI)	Number of items	MWC (IQR)	K_{alpha} (95% CI)
Free-text eligibility criteria forms	142	10 (6-14)	0.19 (0.14-0.24)	150	9 (4.5-13.5)	0.43 (0.37-0.50)
Precoordinated item set	20	3 (2.00-5.25)	0.64 (0.46-0.79)	67	5 (4.00-9.00)	0.78 (0.70-0.86)
Postcoordinated item set	122	10 (6-14)	0.12 (0.08-0.16)	83	11 (8-17.5)	0.16 (0.11-0.21)
Structured quality assurance forms	159	3 (2-4)	0.50 (0.43-0.57)	151[d]	3 (2-4)	0.62 (0.55-0.69)
Precoordinated item set	102	3 (2.00-4.00)	0.72 (0.64-0.80)	116	3 (2.00-4.00)	0.76 (0.69-0.82)
Postcoordinated item set	57	5 (3.00-6.00)	0.12 (0.07-0.15)	33	4 (2.00-6.00)	0.15 (0.07-0.23)

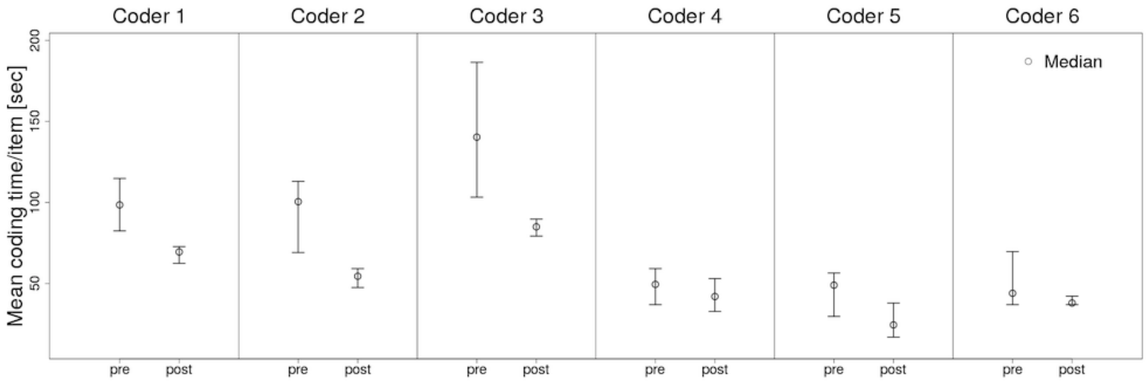
[a] MWC: median word count per item.

[b] IQR: interquartile range.

[c] K_{alpha}: Krippendorff alpha based on 10,000 bootstrapes (95% CI).

[d] Two items could not be coded by any of the coders.

Figure 3. The mean coding times per item for each rater before and after the intervention.



Regarding quality assurance forms, the medians remained the same or slightly increased after the intervention (13, IQR 8.75-23 vs 13.5, IQR 8.25-34.25); this is in accordance to linear regression applied on each of the learning graphs in Figures 4 and 5 (see slope changes before and after intervention). Regarding intercoder time comparisons, the mean intercoder time differences decreased after the intervention in both documentation types (see Figures 4 and 5).

A form item is counted as precoordinated if each coder picked one single (not necessarily the same) Unified Medical Language System code.

Free-Text Eligibility Criteria Versus Structured Quality Assurance Items

Before and after the intervention, structured quality assurance items were associated with significantly higher interrater reliability than free-text eligibility criteria items. However, after the intervention, the difference narrowed down because of a stronger interrater reliability increase in the free-text eligibility criteria items set (before intervention: K_{alpha}=0.17 in eligibility criteria and K_{alpha}=0.50 in quality assurance; after intervention: K_{alpha}=0.42 in eligibility criteria and K_{alpha}=0.62 in quality assurance); see Table 2 for details.

Precoordinated Versus Postcoordinated Concepts

Interrater reliability was significantly higher in precoordinated items versus postcoordinated items before and after intervention regardless of the documentation type. Precoordinated items had an interrater reliability with K_{alpha} ranging from 0.64 to 0.78, and postcoordinated items had an interrater reliability with K_{alpha} ranging from 0.12 to 0.16 (see Table 2 for further details). The coder's decision to pre- or postcoordinate significantly changed after the intervention. The proportion of items that were precoordinated significantly increased both in the eligibility criteria and quality assurance item set (eligibility criteria: OR 4.92, 95% CI 2.78-8.72; quality assurance: OR 1.96, 95% CI 1.19-3.25). Figure 6 provides an example of 2 similar eligibility criteria form items that were coded by different coders to illustrate coding harmonization after intervention. Before intervention, the study inclusion criterion "Written informed consent" was coded the same among coders 2 and 3. Coder 1 used postcoordination with different codes. After intervention, all 3 coders coded the semantically identical inclusion criterion "Informed written consent" using a simplified precoordinated common medical concept, Informed Consent with Unified Medical Language System code C0021430, which was suggested by the code suggestion mechanism and covers sufficiently enough the relevant meaning of inclusion criterion, as shown in Figure 6.

Figure 4. A: The mean coding times for eligibility criteria forms before and after the intervention. B: The number of unique medical concepts each coder has coded on each day.

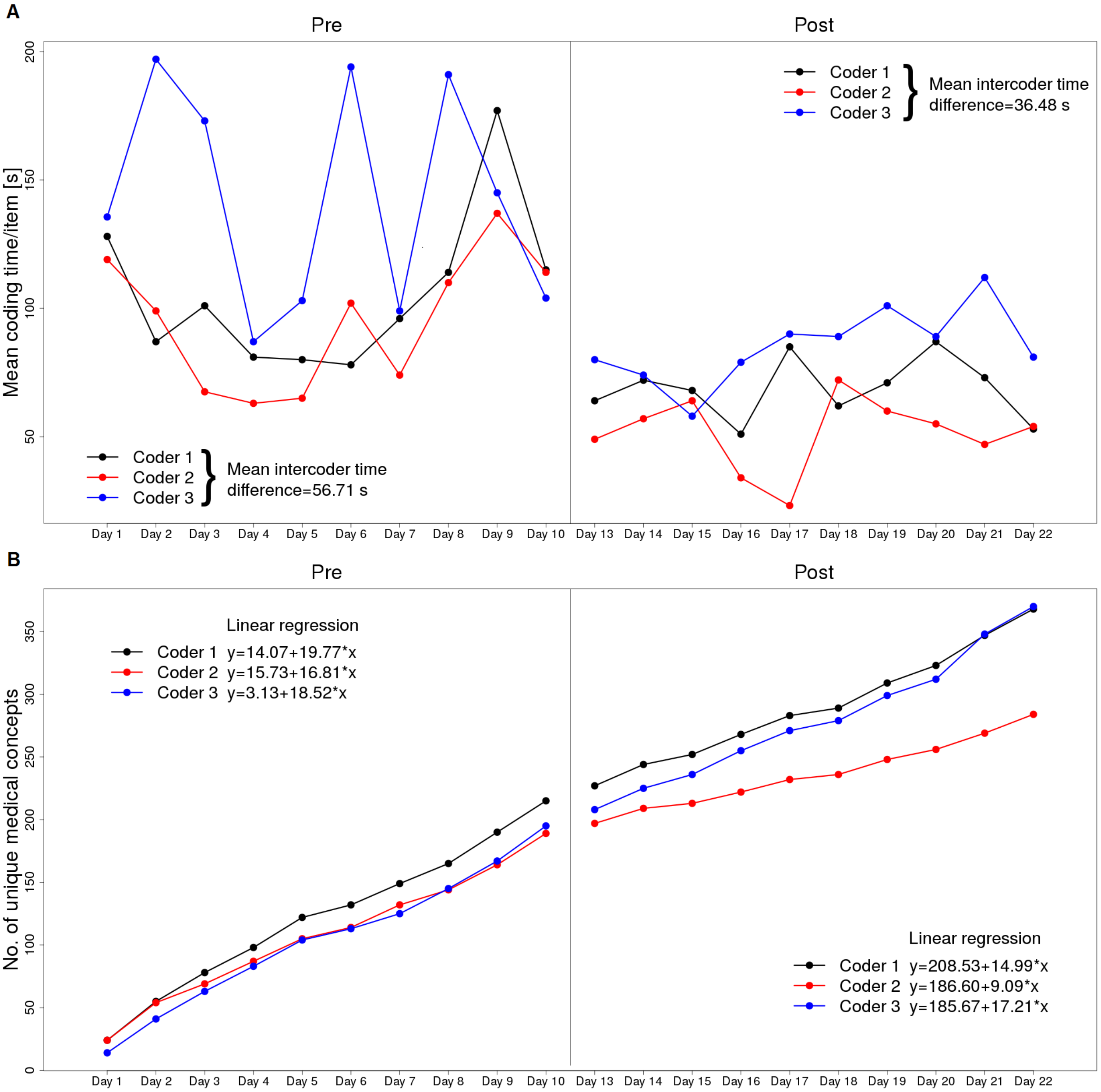


Figure 5. A: The mean coding time per item. B: Daily new unique concepts to code for structured quality assurance forms, analogous to Figure 4.

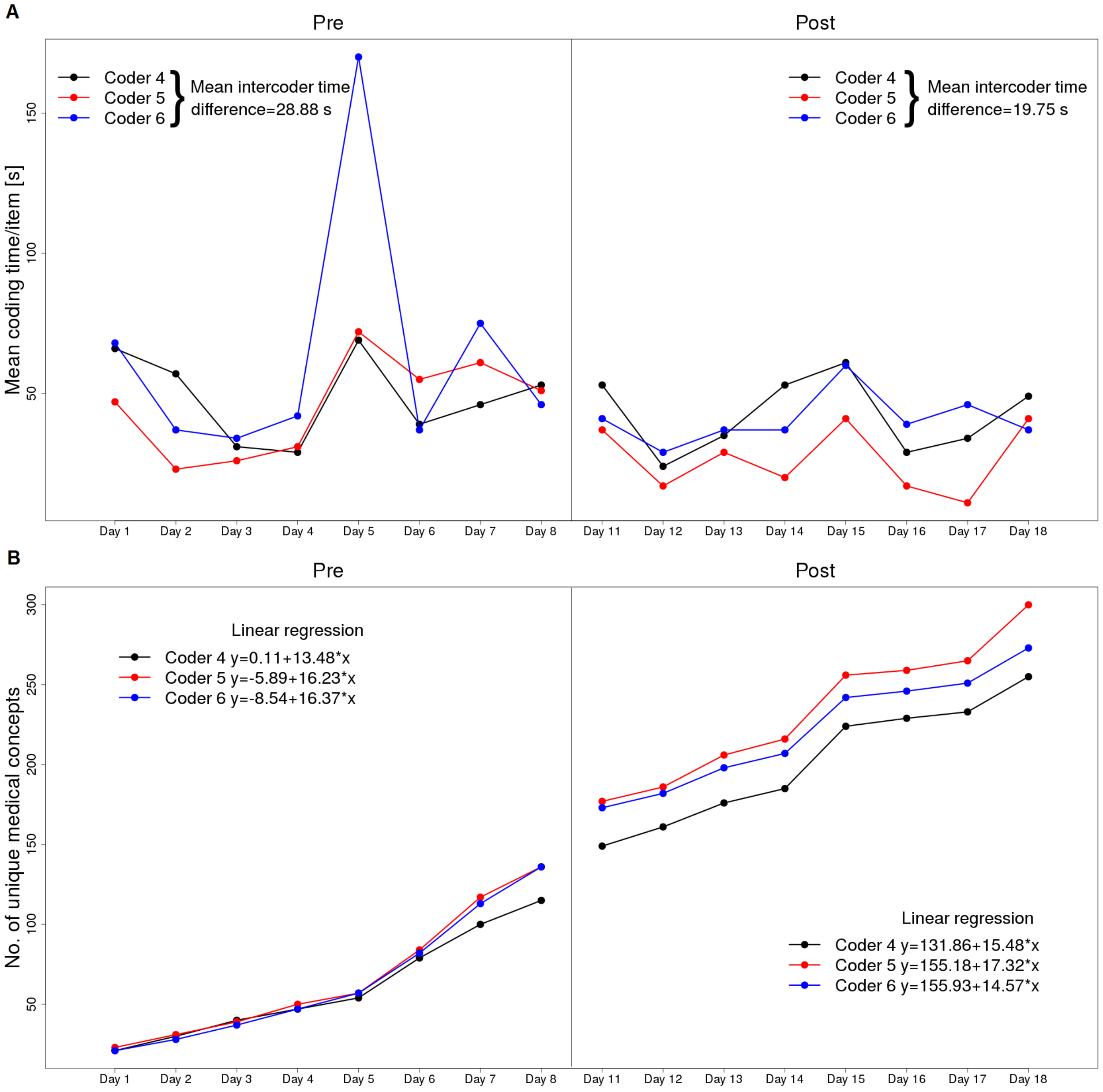
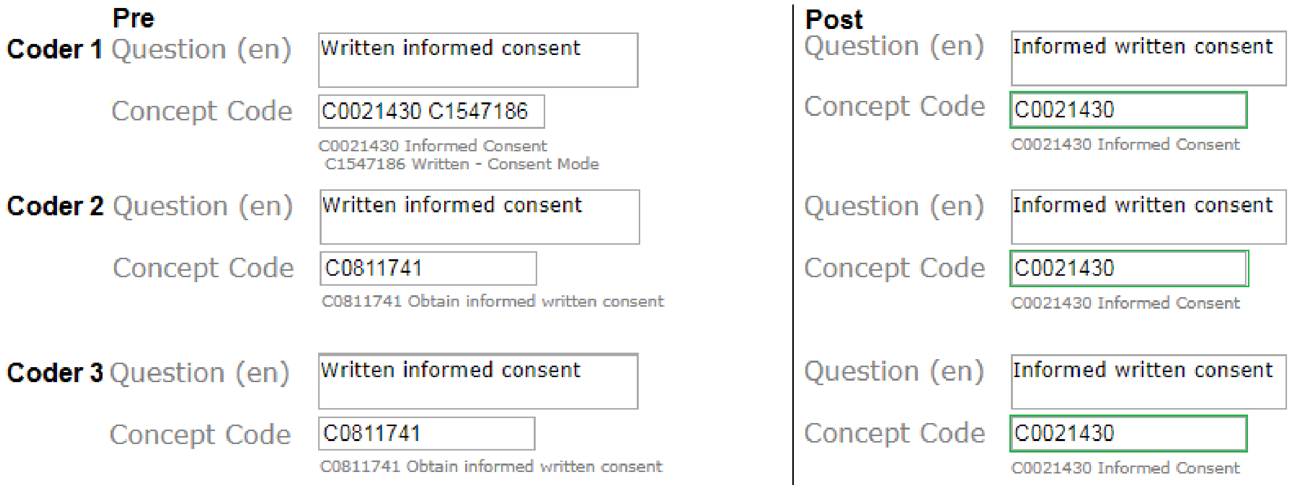


Figure 6. An example of coding harmonization after intervention.



Pre

Coder 1 Question (en) Written informed consent

Concept Code C0021430 C1547186

C0021430 Informed Consent
C1547186 Written - Consent Mode

Coder 2 Question (en) Written informed consent

Concept Code C0811741

C0811741 Obtain informed written consent

Coder 3 Question (en) Written informed consent

Concept Code C0811741

C0811741 Obtain informed written consent

Post

Question (en) Informed written consent

Concept Code C0021430

C0021430 Informed Consent

Question (en) Informed written consent

Concept Code C0021430

C0021430 Informed Consent

Question (en) Informed written consent

Concept Code C0021430

C0021430 Informed Consent

Discussion

Interrater Reliability

Before the intervention, interrater reliability was low or moderate, as reported in related work for other terminologies, such as the emerging clinical reference terminology SNOMED-CT [15,16]. To the best of our knowledge, this is the first study to evaluate the increased interrater reliability through activation of the expert-based code suggestion, which is linked to a large repository of already annotated medical items. Other large terminologies that also suffer from low interrater reliability because of the high abundance of similar or duplicate concepts could benefit from our approach as well.

In this study, the relative increase was more pronounced in free-text eligibility criteria coding. One key observation could explain this phenomenon in the subgroup analysis; owing to the high number of words in an eligibility criterion, coders frequently had chosen postcoordination with rather a word-by-word coding than to search for a common single medical concept to capture the relevant semantics of the whole eligibility criterion. With postcoordination, the probability that another coder chooses the same sequence of codes decreases with each additional postcoordinated code. After the intervention, the coders chose precoordination markedly more often over postcoordination. Although the preference of precoordination is explicitly mentioned as a general rule of thumb as part of the coder training (based on the established coding principles [18]), the code reuse function identifies semantically similar medical text items from a large semantically annotated text corpus and suggests simplified precoordinated coding even for complex free-text items (see example in Figure 6). Because the code suggestion function and the training of coding principles form one coherent unit of the intervention, we did not intend to analyze both parts separately, for example, by further study arms. Therefore, this intervention has to be taken as a whole with respect to reported effects.

According to the K_{alpha} value interpretation by Landis and Koch [33], interrater reliability improved to “moderate agreement” (at least fair agreement regarding 95% CI) and “substantial agreement” (at least moderate agreement regarding 95% CI) in the free-text eligibility criteria set and structured quality assurance set, respectively. The perfect agreement would be required for automatic comparisons based on sole UMLS codes. Therefore, expert-based code review in cases of disagreements might be necessary to rule out false-positive disagreements.

Coding Time

Time measurements indicate slight reductions in coding time in the eligibility criteria subset and a similar coding time in the quality assurance subset. There was a substantial decrease in new concepts in the postinterventional phase for the eligibility criteria set compared with that in the quality assurance set in which the median number of new concepts barely changed; this difference was expectable because the disease-matched eligibility criteria forms were chosen from different clinical trials but they do contain similar inclusion and exclusion criteria. Unlike the eligibility criteria forms, the quality assurance forms stemmed from one source responsible for nationwide quality assurance documentation and, therefore, repetitive data elements among different forms were less common than that in eligibility criteria forms.

Translation Into Natural Language Processing-Based Use Cases

NLP tools that rely on expert-annotated medical text for training can take advantage of this large data repository. As the largest repository of medical data items with semantic codes, it currently consists of >300,000 English medical form items, which are semantically annotated by medical experts with a broad coverage on diverse disease entities [20]. Thus, the meaning of diverse medical text segments, including synonyms and complex clinically relevant concept relations, is machine readable. The study has shown that the use of this large data repository and coding principles improved uniform (=high interrater reliability) coding among different human coding experts. Because NLP pipelines and machine-learning approaches, in general, use expert-annotated text corpora with information coded by different experts, higher interrater reliability would increase the signal-to-noise ratio and, thus, improve semantic classification accuracy in natural free-text. In turn, NLP tools could be more effective in the identification of clinically relevant concepts hidden in clinical notes and corresponding biomedical literature and could be linked to computerized decision support systems for the implementation of evidence-based management strategies at the point of care [34].

Limitations

This study is the first to analyze the effect among different medical coders before and after a training intervention. To the best of our knowledge, a larger set of medical expert coders were never recruited for systematic UMLS intercoder analyses. A larger set of coders with an even larger number of form items to code would have been beneficial to limit the range of dispersion for the reported CIs. However, the strength of the sample size lies in the unprecedented high number of 602 different form items (EC forms pre+post, 292 items; quality assurance forms pre+post, 310 items) covering a broad area of mortality-leading diseases and showing statistical significance in the free-text coding task.

This study had a quasi-experimental design. A randomized controlled design would provide the gold standard to elaborate on the cause and effects of the reported intervention. However, in this study, randomization would have decreased the number of participants assigned to the intervention arm to only 3 participants, whereas the more simplistic pre-post design had the advantage of evaluating all 6 coders with and without intervention and intracoder-related effects (interrater reliability, decision to precoordinate, and coding time) during the full course of the study.

The form content was selected on the basis of the availability and the leading causes of World Health Organization mortality data and, therefore, addresses a broad range of medical concepts in disease entities, which are of high research-interest or under quality assurance. However, the reported effects on interrater reliability might not be generalizable to other clinical fields.

A conceivable success factor of the semiautomatic code suggestion is the underlying annotated text corpus in the MDM portal having oncology as the major disease entity as form content and not mortality-leading diseases such as cardiovascular diseases [20]. Currently, a continuous development of the annotated text corpus is ongoing [35] in many different medical fields and could, therefore, yield a further increase in the interrater reliability in the future.

Conclusions

Coder training and Web-based semiautomatic code suggestion improved interrater reliability in coding medical concepts of diverse mortality-leading disease areas while preserving or even slightly decreasing coding time. Higher interrater reliability represents higher coding uniformity among different medical coders. Consequently, this would lead to a higher signal-to-noise ratio in use cases, which utilize text corpora annotated by multiple coders for semantic analyses.

This study indicates that precoordination in preference to postcoordination and the use of structured data elements in preference to free-text data elements are key drivers for higher interrater reliability. Further development of not only the code suggestion mechanism and use-case specific coder training but also harmonization of codes in the provided medical terminology system are necessary to achieve substantial or almost perfect agreement consistently.

Acknowledgments

This project is funded by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG grant DU 352/11-1).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Coded forms and teaching material.

[ZIP File (Zip Archive), 2MB - jmir_v20i10e274_app1.zip]

References

1. Cios K, William MG. Uniqueness of medical data mining. Artificial Intelligence in Medicine (1-2)? 2002;26:24. [Medline: 12234714]
2. Liu F, Rutherford P, Smoyer-Tomic K, Prichard S, Laplante S. A global overview of renal registries: a systematic review. BMC Nephrol 2015;16(1):31. [Medline: 25886028]
3. Haussig S, Schuler G, Linke A. A. Worldwide TAVI registries: what have we learned? Clin Res Cardiol 2014;103(8):603-612. [Medline: 24648061]
4. Spisla CM, Lundberg CB. Standardization of patient registries for improved data collection and outcome measurement. NI 2012 (2012) 2012;2012:391 [FREE Full text] [Medline: 24199127]
5. Dugas M, Meidt A, Neuhaus P, Storck M, Varghese J. ODMedit: uniform semantic annotation for data integration in medicine based on a public metadata repository. BMC Med Res Methodol 2016 Dec 01;16:65 [FREE Full text] [doi: 10.1186/s12874-016-0164-9] [Medline: 27245222]
6. Dugas M. Missing semantic annotation in databases. The root cause for data integration and migration problems in information systems. Methods Inf Med 2014;53(6):516-517.
7. Varghese J, Fujarski M, Hegselmann S, Neuhaus P, Dugas M. CDEGenerator: an online platform to learn from existing data models to build model registries. Clin Epidemiol 2018;10:961-970. [doi: 10.3233/978-1-61499-678-1-282]
8. Tseytlin E, Mitchell K, Legowski E, Corrigan J, Chavan G, Jacobson RS. NOBLE - Flexible concept recognition for large-scale biomedical natural language processing. BMC Bioinformatics 2016 Jan 14;17:32 [FREE Full text] [doi: 10.1186/s12859-015-0871-y] [Medline: 26763894]
9. Savova G, Masanz J, Ogren P, Zheng J, Sohn S, Kipper-Schuler K, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc 2010;17(5):507-513 [FREE Full text] [doi: 10.1136/jamia.2009.001560] [Medline: 20819853]
10. Becker M, Böckmann B. Extraction of UMLS Concepts Using Apache cTAKES for German Language. In: Stud Health Technol Inform. 2016 Presented at: Health Informatics meets eHealth conference; 2016; Vienna p. 6. [doi: 10.3233/978-1-61499-645-3-71]
11. Divita G, Zeng Q, Gundlapalli A, Duvall S, Nebeker J, Samore M. Sophia: A Expedient UMLS Concept Extraction Annotator. In: AMIA Annu Symp Proc. 2014 Presented at: AMIA Annu Symp Proc; 2014; Washington D.C p. 467-476.
12. Aronson A, Lang FM. An overview of MetaMap: historical perspective and recent advances. J Am Med Inform Assoc 2010;17(3):229-236.

13. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004;32(Database issue):267-270.
14. UMLS Similarity Web Interface. URL: http://maraca.d.umn.edu/cgi-bin/umls_similarity/umls_similarity.cgi [accessed 2017-12-07] [WebCite Cache ID 6vXQD6e8A]
15. Andrews J, Richesson R, Krischer J. Variation of SNOMED CT coding of clinical research concepts among coding experts. J Am Med Inform Assoc 2007;14(4):497-506 [FREE Full text] [doi: 10.1197/jamia.M2372] [Medline: 17460128]
16. Chiang MF, Hwang JC, Yu AC, Casper DS, Cimino JJ, Starren JB. Reliability of SNOMED-CT coding by three physicians using two terminology browsers. In: AMIA Annu Symp Proc. 2006 Presented at: AMIA Annu Symp Proc ?5; 2006; Washington, D.C p. 131-135 URL: http://europepmc.org/abstract/MED/17238317
17. Rothschild AS, Lehmann HP, Hripcsak G. Inter-rater agreement in physician-coded problem lists. In: AMIA Annu Symp Proc. 2005 Presented at: AMIA Annu Symp Proc; 2005; Washington, D.C p. 644.
18. Varghese J, Dugas M. Frequency analysis of medical concepts in clinical trials and their coverage in MeSH and SNOMED-CT. Methods Inf Med 2015;54(1):83-92.
19. Benson T. Principles of health interoperability HL7 and SNOMED. Second Edition. Dordrecht: Springer; 2012.
20. Dugas M, Neuhaus P, Meidt A, Doods J, Storck M, Bruland P, et al. Portal of medical data models: information infrastructure for medical research and healthcare. Database (Oxford) 2016:1.
21. WHO. WebCite. WHO 2015 Global Health estimate data URL: http://www.who.int/healthinfo/global_burden_disease/en [WebCite Cache ID 6vU8P0czP]
22. Institute for Applied Quality Improvement and Research in Health Care (AQUA). WebCite URL: https://www.aqua-institut.de/ [accessed 2017-12-05] [WebCite Cache ID 6vU8S4tp0]
23. Varghese J, Sünninghausen S, Dugas M. Standardized Cardiovascular Quality Assurance Forms with Multilingual Support, UMLS Coding and Medical Concept Analyses. Stud Health Technol Inform 2015;Volume 216: MEDINFO 2015: eHealth-enabled Health:837-841. [Medline: 26262169]
24. Varghese J, Sünninghausen S, Dugas M. Standardized quality assurance forms for organ transplantations with multilingual support, open access and UMLS coding. Stud Health Technol Inform 2015;Volume 212: eHealth2015 – Health Informatics Meets eHealth:22. [Medline: 26063252]
25. The Clinical Data Interchange Standards Consortium. WebCite URL: https://www.cdisc.org/ [accessed 2017-12-05] [WebCite Cache ID 6vTx9OyHI]
26. Scott W. Reliability of Content Analysis: The Case of Nominal Scale Coding. Public Opinion Quarterly 1955;19(3):321.
27. Cohen J. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 1960;20(1):37-46. [doi: 10.1177/001316446002000104]
28. Thompson W, Walter S. A reappraisal of the kappa coefficient. J Clin Epidemiol 1988;41(10):949-958. [Medline: 3057117]
29. Kraemer H, Bloch D. Kappa coefficients in epidemiology: An appraisal of a reappraisal. J Clin Epidemiol 1988;41(10):959-968.
30. Feinstein A, Cicchetti D. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol 1990;43(6):543-549. [Medline: 2348207]
31. Zapf A, Castell S, Morawietz L, Karch A. Measuring inter-rater reliability for nominal data - which coefficients and confidence intervals are appropriate? BMC Med Res Methodol 2016 Dec 05;16:93 [FREE Full text] [doi: 10.1186/s12874-016-0200-9] [Medline: 27495131]
32. Krippendorff K. Estimating the Reliability, Systematic Error and Random Error of Interval Data. Educational and Psychological Measurement 1970;30(1):61-70.
33. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
34. Afzal N, Mallipeddi V, Sohn S, Liu H, Chaudhry R, Scott C, et al. Natural language processing of clinical notes for identification of critical limb ischemia. Int J Med Inform 2018 Dec;111:83-89 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.12.024] [Medline: 29425639]
35. Medical-Data-Models Portal. WebCite URL: https://medical-data-models.org/ [accessed 2017-12-05] [WebCite Cache ID 6vU8QSKEB]

Abbreviations

IQR: interquartile range
UMLS: Unified Medical Language System
K_{alpha}: Krippendorff alpha
MDM: Medical Data Models
NLP: Natural Language Processing
ODM: operational data model

Edited by G Eysenbach; submitted 13.12.17; peer-reviewed by I Adeleke, J Bian, H Oh; comments to author 26.04.18; revised version received 03.05.18; accepted 28.06.18; published 15.10.18.

Please cite as:
Varghese J, Sandmann S, Dugas M
Web-Based Information Infrastructure Increases the Interrater Reliability of Medical Coders: Quasi-Experimental Study
J Med Internet Res 2018;20(10):e274
URL: http://www.jmir.org/2018/10/e274/
doi:10.2196/jmir.9644
PMID:30322834

Original Paper

Complaint-Directed Mini-Interventions for Depressive Symptoms: A Health Economic Evaluation of Unguided Web-Based Self-Help Interventions Based on a Randomized Controlled Trial

Ben FM Wijnen[1,2], PhD; Suzanne Lokman[3], MSc; Stephanie Leone[3], PhD; Silvia MAA Evers[1,2], PhD; Filip Smit[1,4,5], PhD

[1]Centre for Economic Evaluation, Trimbos Institute, Netherlands Institute of Mental Health and Addiction, Utrecht, Netherlands

[2]Department of Health Services Research, Care and Public Health Research Institute School for Public Health and Primary Care, Maastricht University, Maastricht, Netherlands

[3]Department of Public Mental Health, Trimbos Institute, Netherlands Institute of Mental Health and Addiction, Utrecht, Netherlands

[4]Department of Clinical Psychology, Faculty of Behavioural and Movement Sciences, VU University, Amsterdam, Netherlands

[5]Amsterdam Public Health Research Institute, Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, Netherlands

Corresponding Author:
Ben FM Wijnen, PhD
Centre for Economic Evaluation
Trimbos Institute
Netherlands Institute of Mental Health and Addiction
Da Costakade 45
Utrecht, 3521 VS
Netherlands
Phone: 31 030 297 1100
Email: b.wijnen@maastrichtuniversity.nl

Abstract

Background: Depression prevention and early intervention have become a top priority in the Netherlands, but with considerable room for improvement. To address this, Web-based complaint-directed mini-interventions (CDMIs) were developed. These brief and low-threshold interventions focus on psychological stress, sleep problems, and worry, because these complaints are highly prevalent, are demonstrably associated with depression, and have substantial economic impact.

Objective: The objective of this economic evaluation was to examine the added value of Web-based, unguided, self-help CDMIs compared with a wait-listed control group with unrestricted access to usual care from both a societal and a health care perspective.

Methods: This health economic evaluation was embedded in a randomized controlled trial. The study entailed 2 arms, in which 3 Web-based CDMIs were compared with a no-intervention waiting-list control group (which received the intervention after 3 months). We conducted measurements at baseline, and at 3 and 6 months. The primary outcome was the rate of responders to treatment on depressive symptoms as measured by the Inventory of Depressive Symptomatology Self-Report (IDS-SR). We estimated change in quality of life by calculating effect sizes (Cohen *d*) for individual pre- and posttreatment IDS-SR scores using a conversion factor to map a change in standardized effect size onto a corresponding change in utility. We calculated incremental cost-effectiveness ratios using bootstraps (5000 times) of seemingly unrelated regression equations and constructed cost-effectiveness acceptability curves for the costs per quality-adjusted life-year (QALY) gained.

Results: Of 329 study participants, we randomly assigned 165 to the CDMI group. At 3 months, the rate of responders to treatment was 13.9% (23/165) in the CDMI group and 7.3% (12/164) in the control group. At 3 months, participants in the CDMI group gained 0.15 QALYs compared with baseline, whereas participants in the control group gained 0.03 QALYs. Average total costs per patient at 3 months were €2094 for the CDMI group and €2230 for the control group (excluding baseline costs). Bootstrapped seemingly unrelated regression equations models resulted in a dominant incremental cost-effectiveness ratio (ie, lower costs and a higher rate of responders to treatment) for the CDMI group compared with the control group at 3 months, with the same result for the costs per QALY gained. Various sensitivity analyses attested to the robustness of the findings of the main analysis.

Conclusions: Brief and low-threshold Web-based, unguided, self-help CDMIs have the potential to be a cost-effective addition to usual care for adults with mild to moderate depressive symptoms. The CDMIs improved health status, while reducing participant health care costs, and hence dominated the care-as-usual control condition. As intervention costs were relatively low, and the internet is readily available in the Western world, we believe CDMIs can be easily implemented on a large scale.

Trial Registration: Netherlands Trial Register NTR4612; http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=4612 (Archived by WebCite at http://www.webcitation.org/6n4PVYddM)

(J Med Internet Res 2018;20(10):e10455) doi:10.2196/10455

KEYWORDS

prevention; depression; internet-based intervention; economic evaluation; quality of life; cost-effectiveness analysis; cost-utility analysis; early medical intervention; cost-benefit analysis

Introduction

The Burden of Depression

Globally, more than 300 million people from different age groups have depression [1]. Likewise, depression is the most prevalent psychological disorder in the Netherlands. In 2017, about 550,500 people aged 18 to 65 years had a depression disorder [2]. In addition, slightly more than 1 million people (1,006,700) had subclinical depression (ie, people with the core symptoms of depression otherwise not fulfilling the diagnostic criteria of major depression) [3]. However, depression not only causes individual suffering and loss of quality of life [4], but is also associated with economic costs as a result of health care utilization and reduced productivity owing to absenteeism and lesser efficiency while at work [5]. In 2011, the cost of treatments for depression was €1.6 billion in the Netherlands [6]. This corresponds to 1.8% of the total expenditure on health care. Moreover, a study by Greenberg et al estimated that 48% to 50% of the total costs of depression were related to workplace costs [7].

Given the high prevalence and the chronic character of the condition, prevention and early intervention is important, especially as current psychological and pharmacological interventions have been shown to only moderately reduce the burden of depression [2,8].

Web-Based Complaint-Directed Mini-Interventions

In the past decades, depression prevention and early intervention has become a top priority in the Netherlands. However, despite evidence that such programs can be effective, there is still considerable room for improvement. Specifically, there is a need for novel interventions that are easily accessible, cheap, and, importantly, suitable for high-risk populations (eg, people with a low socioeconomic status), as the reach of preventive or early mental health interventions is far from optimal among these populations. It is also important that interventions are able to encourage self-management and can be implemented against limited costs. With this in mind, the Web-based complaint-directed mini-interventions (CDMIs) were developed [9]. The unique feature of the CDMIs as an approach for depression is that they were developed by taking into account that symptoms preceding or underlying depression may not be disorder specific (eg, worry) and may also vary by individual, which is in line with recent symptom network and transdiagnostic approaches to mental disorders [10-12]. These brief and low-threshold interventions focus on psychological stress, sleep problems, and worry, because these complaints are highly prevalent, are demonstrably associated with depression, have a substantial economic impact, and are frequently presented to the general practitioner [13-15]. Hence, CDMIs focus on tackling complaints (symptoms) rather than on disorders, which allows each individual to choose the complaint(s) they want to focus on according to their needs. This feature may be advantageous, as it may provide a better fit between what an individual needs and the intervention. In total, 3 different CDMIs were developed: Sleep better, Stress less, and Worry less.

Although the impact of interventions on individual and population health is vital, it is also important to determine the role of different interventions in contributing to other socially desirable goals, such as reducing societal (health care) costs. In addition, due to scarcity of resources for and rising costs of the health care system, economic evidence is becoming increasingly important for decision makers and regulatory bodies. Health economic evaluations aim to provide economic evidence on the costs and effects of (new) health care interventions. In a health economic evaluation, both the costs and effects of 2 (or more) alternative treatments or interventions are compared in a systematic manner. In this way, it is possible to examine which alternative is most efficient and hence provides the best value for money [16]. The use of economic evaluations is required by the National Health Care Institute in the Netherlands [17] and by the National Institute for Health and Care Excellence in the United Kingdom [18]. Treating people in a relatively accessible way in an early stage may prevent them from developing more serious mental disorders (eg, chronic depression), and possibly prevent them from needing high-cost mental health care.

Web-based interventions for depressive symptoms have been shown to be cost effective in the past. For example, McCrone et al assessed the cost effectiveness of computer-delivered cognitive behavioral therapy and concluded that it has a high probability of being cost effective [19]. Moreover, Warmerdam et al evaluated the cost effectiveness of internet-based cognitive behavioral therapy and internet-based problem-solving therapy and concluded that both have a high probability of being cost effective [20]. Both studies, however, examined interventions that were relatively extensive compared with the CDMIs.

Objective

The aim of this economic evaluation was to examine the added value of the Web-based unguided self-help CDMIs as compared with a wait-listed control group with unrestricted access to usual care from both a societal and a health care perspective. We conducted the health economic evaluation in a sample of adults with mild to moderate depressive symptoms from a societal perspective. In addition, we conducted the analysis from an employers' perspective to determine the cost and effects associated with CDMIs specifically related to productivity losses in the subsample of people with a paid job.

Methods

Parent Randomized Controlled Trial

This economic evaluation was embedded in a randomized controlled trial (see Lokman et al [21]). The study entailed 2 arms in which we compared 3 Web-based CDMIs with a no-intervention waiting-list control group. We used stratified block randomization to ensure that participants were equally distributed over the 3 different CDMIs (ie, Sleep better, Stress less, or Worry less) and 2 levels of education (high: higher vocational or university-level education; or low educated). Measurements were conducted at baseline, and at 3 and 6 months' follow-up.

Study Population and Recruitment

Patients were included if they fulfilled the following criteria: (1) at least 18 years of age, (2) access to a computer with an internet connection, (3) sufficient proficiency of the Dutch language, (4) adequate computer skills to participate in the training, and (5) mild to moderate depressive symptoms defined as a score of 14 to 38 on the Inventory of Depressive Symptomatology Self-Report (IDS-SR) [22]. These IDS-SR cutoff scores imply that the CDMIs were used for indicated prevention in subclinical depression and early intervention in mild depression. Participants were excluded if they had suicidal thoughts or plans as measured with item 18 of the IDS-SR. These participants were referred to contact their general practitioner or an anonymous online platform for people with suicidal thoughts. The required sample size for this study to achieve a power of 80% was estimated to be 292 (146 per condition), based on achieving an effect size of 0.33 with a power of .80 and a 2-tailed test with alpha=.05.

Participants were recruited from June 2014 to January 2015 via open recruitment (ie, through relevant websites, messages on social media, and messages in digital newsletters of the Trimbos Institute, Utrecht, the Netherlands). Next, people interested in participation were referred to a special study website where they were given more information about the study and could register to take part in the study by completing a written or an online informed consent form including their name and email address. Applicants were requested to complete the first part of the self-report online baseline questionnaire, which acted as a screening instrument and consisted of the IDS-SR and questions about age, internet access, and computer skills. Eligible participants received the second part of the online baseline questionnaire. To be able to conduct the stratified block randomization, we asked applicants which CDMI they would want to take part in: Sleep better, Stress less, or Worry less. The study was approved by the Medical Ethics Committee of the University Medical Center Utrecht and is registered in the Netherlands Trial Register (NTR): NTR4612. More details can be found in Lokman et al [21].

Intervention

The 3 CDMIs are unguided, Web-based, self-help interventions to prevent or reduce depressive complaints. As noted in the introduction, symptoms preceding or underlying (a developing) depression may not be disorder specific (eg, worry) and may also vary by individual. Therefore, the CDMIs were developed with a focus on specific complaints rather than being a program targeting a subclinical disorder. This allows each individual to choose the complaint(s) they want to focus on based on their needs. Thus, the CDMIs were developed taking into account that symptoms of depression or a developing depression varies by individual while still serving as an overall approach to combat depressive symptoms.

The CDMIs are therefore complaint focused rather than disorder focused. The content of the CDMIs is largely based on cognitive behavioral techniques but also incorporates elements from solution-focused therapy, mindfulness, and positive psychology. The CDMIs are made up of 3 to 4 modules, with each module consisting of 4 to 6 exercises. Some modules, such as relaxation, are relevant for all 3 complaints (sleep, stress, and worry) and are, therefore, part of all 3 CDMIs. Fixed elements in every CDMI are a home page, a diary, a list of the user's favorite exercises, an exercise book, a to-do list, and a library. Users were free to choose between the modules and exercises and could work independently through the CDMI, without supervision. Participants received a reminder if they did not log in to the CDMI within 1 week after registration. Participants were advised to spend 2 to 3 hours a week on the CDMI for a period of at least 4 weeks. Multimedia Appendix 1 (adapted from Lokman et al [21]) contains a more detailed description of the CDMIs (including screenshots). All CDMIs were developed by the Trimbos Institute (Netherlands Institute for Mental Health and Addiction).

Control Group

Participants randomly assigned to the control group were placed on a waiting list for 3 months with unrestricted care as usual. They were provided access to the CDMI of their choice afterward.

Outcomes

The primary outcome of the study was depressive symptoms as measured by the IDS-SR [22]. The IDS-SR consists of 30 items relating to the last 7 days that cover 9 diagnostic symptom domains used to characterize a major depressive episode, as well as items to define melancholic and atypical symptom features, commonly associated symptoms (eg, irritability, anxiety), and features of endogenous symptoms. We chose the IDS-SR as the primary outcome because the study focused on adults with mild to moderate depressive symptoms, and we hypothesized a greater reduction in depressive complaints for the participants using the Web-based CDMIs. Items are scored on a 4-point Likert scale and can be summed to obtain a total score. Scores range from 0 to 84, with higher scores indicating greater depressive symptom severity. For the cost-effectiveness analysis, we used the rate of responders to treatment as the central clinical end term, arbitrarily chosen to ease interpretation of the cost-effectiveness analysis (ie, additional costs per responder instead of additional costs per percentage improvement on IDS-SR scores). Rate of responders to treatment was defined as having a decrease on the IDS-SR scale by 50% (or more) compared with baseline.

Then, we estimated a change in quality of life by calculating effect sizes (Cohen *d*) for individual pre- and posttreatment IDS-SR scores (ie, Cohen *d* = (IDS-SR[TX]–IDS-SR[T0]) / standard deviation of IDS-SR[T0]), where T0 is baseline and TX values are the assessments at the 3- and 6-month follow-ups. Next, we used the conversion factor of Sanderson and colleagues to map a change in standardized effect size onto a corresponding change in utility for people with depression [23]. This conversion factor entails the average difference in utility that is associated with a difference of 1 effect size. Utilities represent the value of a particular health state on a scale anchored at 0 and 1, in which 0 means death and 1 means perfect health. We then used the utilities to calculate quality-adjusted life-years (QALYs) by multiplying them by the time spent in that particular health state [16].

Costs

We used the Dutch guidelines for economic evaluations [24] and the Consolidated Health Economic Evaluation Reporting Standards [25] to conduct the economic evaluation and to report the outcomes. Conforming to the Dutch guidelines, we took a societal perspective, in which all relevant costs for society should be taken into consideration. Based on this guidance, these costs entail patient and family costs (eg, travel costs, home care, informal care) and productivity losses (ie, presenteeism and absenteeism from work).

We measured (health care) resource use, costs for patient and family, and productivity losses using the Trimbos/iMTA Questionnaire for Costs Associated With Psychiatric Illness [26].

We distinguished 4 cost categories: intervention costs, health care sector costs, costs for patient and family, and costs owing to productivity losses. Intervention costs were based on the total number of accounts per year, the costs for hosting and updating the website, and costs for support by a helpdesk that could be reached by phone. To value cost items, we used standardized cost prices from the Dutch manual for costing [24]. If those were not available, we used mean cost prices from providers. To determine the costs of drugs, we asked participants for how many days they had used drugs for depression, problems sleeping, or anxiety. Next, we assigned monetary values based on an average cost price per day for depression, sleep, or anxiety disorders separately as determined by average cost per day (using data from the Dutch Healthcare Institute [27]) for a selection of the most prescribed drugs for each category in combination with their recommended daily dose (using data from the Dutch Healthcare Institute [28]).

As recommended in the Dutch guidelines, productivity losses were estimated using the friction cost approach. The friction cost approach entails the calculation of productivity losses only during a prespecified friction period (85 days according to the Dutch guidelines) [24]. This period is supposed to be the time until another worker from the pool of unemployed has fully replaced the individual who is absent due to an illness. We considered both absenteeism (absence from work due to sickness) and presenteeism (reduced productivity). Furthermore, we valued patients’ time and informal care using the proxy good method, using the average hourly wage of domestic help as a proxy.

All costs were indexed for the year 2016. Given the follow-up of the study, no discounting was performed.

Analyses

We carried out all analyses while adhering to the intent-to-treat principle; that is, we analyzed all participants as randomized provided that their baseline data were complete. For these analyses, we imputed missing values at the follow-up measurements using multiple imputation (5 times). Imputation for total costs and IDS-SR scores was based on age, sex, group, baseline IDS-SR score, baseline Jenkins Sleep Evaluation Questionnaire (JSEQ) score, baseline Perceived Stress Scale (PSS) score, baseline Generalized Anxiety Disorder 7-item (GAD-7) scale score, and baseline health care, patient and family, and productivity costs (for cost data only). To account for nonnormality of the data, we used predictive mean matching, in which real observed values from similar cases are imputed instead of imputing regression estimates [29,30].

We used rate of responders to treatment, as determined by at least a 50% decrease in IDS-SR scores, to calculate an incremental cost-effectiveness ratio (ICER) by dividing the difference in costs by the difference in rate of responders to treatment between both groups. This resulted in the additional costs per extra responder. In addition, we used QALY estimates to calculate the incremental cost-utility ratio (ICUR) by dividing the difference in costs by the difference in QALY between both groups. This way, the ICUR represents the additional costs per QALY gained. To investigate the uncertainty around the ICER and ICUR, we used nonparametric bootstrapping (5000 times). Bootstrapping is a nonparametric way to repeatedly simulate an analysis by resampling, with replacement, from the observed data [31]. We bootstrapped (5000 times) seemingly unrelated regression equations (SUREs) to allow for correlated residuals of the cost and utility equations and to account for baseline differences in productivity costs. Next, we constructed cost-effectiveness acceptability curves for the costs per QALY gained, in which the likelihood that the CDMIs are cost effective is presented given several willingness-to-pay ceilings. A report from the Dutch Council for Public Health and Health Care provided guidance on the ceiling ratios for a QALY for diseases defined by their disability weight. Based on this report, the ceiling ratio can be roughly estimated to be €20,000 to €80,000 per QALY depending on the severity of the disease or disorder [32], and since we were looking at subclinical and mild manifestations of depression, this would put the willingness-to-pay ceiling at €20,000 per QALY gained.

We present comparative results for 3 months’ follow-up, as the control group received the CDMIs after 3 months of follow-up, which hampers interpretation at 6 months’ follow-up. We used results at 6 months only to check whether effects at 3 months were sustained. All analyses were carried out using Stata 14 (StataCorp LLC).

Sensitivity Analyses

We conducted sensitivity analyses to assess the impact of certain assumptions on the results presented in the base case for 3 months' follow-up (as this was the end of the comparative phase). First, we performed a subgroup analysis of only participants with a paid job to determine the effect of CDMIs on productivity losses (due to both absenteeism and presenteeism) in relation to the intervention costs (ie, investment costs from an employers' perspective). Second, we conducted an analysis without baseline adjustments (ie, as the base case analyses were based on baseline-adjusted estimates) to determine the impact of these adjustments on the ICERs and ICURS. Third, we performed analyses from a health care perspective, which excluded all patient and family costs and productivity losses. We did this because the health care perspective is a dominant perspective in health economics and is recommended as the main perspective in certain countries, for example, in the United Kingdom by The National Institute for Health and Care Excellence [33]). Fourth, although opportunity costs for participants (their time spent using the CDMIs) should not be included according to the Dutch guidelines, we conducted an additional sensitivity analysis in which we incorporated these costs in the intervention costs of these self-help CDMIs to see whether this would lead to different conclusions. Fifth, given the relatively large dropout at T1 (3 months), we conducted additional sensitivity analyses in which we used 2 different imputation techniques. First, we used regression-based imputation, in which we checked which predictors were significantly associated with dropout at T1 and which predictors were significantly associated with the primary outcome. We then used these variables in a linear regression to impute missing values [34]. Second, we used a simple imputation based on last observation carried forward per patient. Using different imputation techniques, we were able to determine whether the use of a different approach would have affected our base case results.

Results

Sample at Baseline

In total, we included 329 participants in the study, of whom we randomly assigned 165 to the CDMI condition and 164 to the (waiting-list) control condition. Participants in the CDMI group were distributed among the 3 interventions as follows: Sleep better, n=59 (35.7%); Stress less, n=45 (27.3%); and Worry less, n=61 (37.0%). During the 3-month intervention period, the participants logged in a median of 3 times (range 0-166, interquartile range 5). After 3 months, participants in the waiting list were allowed access to the interventions and were distributed in a similar pattern: Sleep better, n=60 (36.7%); Stress less, n=43 (26.2%); and Worry less, n=61 (37.2%). Table 1 presents the baseline characteristics of all participants.

Loss to Follow-Up

At 3 months, 68 participants were lost to follow-up in the CDMI group (41.2%) and 24 participants in the control group (14.6%), which was statistically significant (χ^2_1=28.8, P<.001). Hence, we performed a sensitivity analyses using covariates significantly associated with dropout at 3 months. At 6 months, 97 participants were lost to follow-up in the CDMI group (58.8%) and 82 in the control group (50.0%) cumulatively. This resulted in 68 participants with complete follow-up in the CDMI group and 82 in the control group.

Clinical Outcomes

At 3 months' follow-up, the rate of responders to treatment was 13.9% (23/165) in the CDMI group and 7.3% (12/164) in the control group. At 6 months' follow-up rates of responders to treatment were 18.8% (31/165) and 11.6% (19/164), respectively. When looking at quality of life, participants in the CDMI group gained 0.15 QALY at both 3 and 6 months' follow-up compared with baseline, whereas participants in the control group gained 0.03 QALY at 3 months' follow-up and 0.16 QALY at 6 months' follow-up (see Table 2 [23]). A more detailed analysis on the clinical outcomes can be found in Lokman et al [21]. In short, Lokman et al demonstrated a significant reduction in depressive symptoms for participants in the intervention group compared with participants in the waiting-list control group after 3 months' follow-up. Furthermore, significant effects were observed for sleep problems, worry, anxiety, and well-being [21].

Costs

At baseline, productivity losses were higher in the CDMI group (see Table 3). Hence, we adjusted bootstrapped SURE models to correct for this baseline difference.

Average total costs per patient (intervention costs, health care costs, patient and family costs, and productivity losses) during the 3-month follow-up were €2094 for the CDMI group and €2230 for the control group (excluding baseline costs; Table 4). At 6 months' follow-up, total costs were €3643 for the CDMI group and €3534 for the control group. No large cost differences were demonstrated between the 2 groups. We estimated intervention costs based on the yearly number of accounts (5000) at €3.90 per participant. When including the participants' time costs (opportunity costs), intervention costs were €52.9 per participant.

Incremental Cost-Effectiveness Ratios

Bootstrapped SURE models, in which we made a baseline adjustment regarding productivity losses, resulted in a dominant ICER (ie, lower costs and a higher rate of responders to treatment) for the CDMI group compared with the control group at 3 months' follow-up (see Figure 1). Likewise, the ICURs in terms of QALY gain were dominant (ie, lower costs and increases in utility) when compared with the control group at 3 months' follow-up (see Figure 2).

Table 1. Demographic characteristics and clinical information on participants in the complaint-directed mini-intervention (CDMI) and control groups at baseline (N=329).

Characteristic	CDMI group (n=165)	Control group (n=164)
Age (years), mean (SD), range	42.85 (12.83), 18-76	43.65 (13.05), 18-81
Sex, n (%)		
Female	122 (73.9)	127 (77.4)
Male	43 (26.1)	37 (22.6)
Marital status, n (%)		
Single	83 (50.3)	84 (51.2)
Living with partner	82 (49.7)	80 (48.8)
Nationality, n (%)		
Dutch	2 (1.8)	4 (2.4)
Other	160 (97.6)	163 (98.2)
Living arrangement, n (%)		
Alone	40 (24.2)	39 (23.8)
With other	125 (75.8)	125 (76.2)
Education level, n (%)		
Low	50 (30.3)	48 (29.3)
High	115 (69.7)	116 (70.7)
Employment, n (%)		
Paid	116 (70.3)	117 (71.3)
Unpaid	49 (29.7)	47 (28.7)
Duration of complaints (years), n (%)		
<1	59 (35.8)	63 (38.4)
≥1	106 (64.2)	101 (61.6)
Severity of complaints, n (%)		
Low	68 (41.2)	83 (50.6)
High	97 (58.8)	81 (49.4)
Evaluation scores, mean (SD)		
Sleep (Jenkins Sleep Evaluation Questionnaire)	11.61 (5.42)	11.21 (5.34)
Stress (Perceived Stress Scale)	21.82 (5.86)	21.48 (5.37)
Worry (Penn State Worry Questionnaire)	37.76 (9.26)	38.28 (9.61)
Anxiety (Generalized Anxiety Disorder 7-item)	10.09 (4.16)	10.04 (3.73)

Table 2. Rates of responders to treatment and utilities at baseline, and 3- and 6-month follow-ups by group (N=329) for the complaint-directed mini-intervention (CDMI) and control groups.

Measure	CDMI group (n=165)			Control group (n=164)		
	Baseline	3 months	6 months	Baseline	3 months	6 months
Rates of responders to treatment based on IDS-SR[a], n (%)	N/A[b]	23 (13.9)	31 (18.8)	N/A	12 (7.3)	19 (11.6)
Utilities and quality-adjusted life-years[c] gained, mean (SD)	N/A	0.15 (0.02)	0.15 (0.02)	N/A	0.03 (0.02)	0.16 (0.01)

[a] IDS-SR: Inventory of Depressive Symptomatology Self-Report.

[b] N/A: not applicable.

[c] Calculated using a translation factor to transform differences in effect size to changes in utility for people with depression developed by Sanderson et al [23].

Table 3. Average per-patient baseline costs for the complaint-directed mini-intervention (CDMI) and control groups[a].

Cost type	CDMI group (n=165)		Control group (n=164)	
	Mean cost (€)	95% CI	Mean cost (€)	95% CI
Total health care costs	197	162-232	243	146-340
Total patient and family costs	87	16-159	53	15-91
Total productivity losses	729	520-937	582	387-778

[a] 1 month; indexed for the year 2016.

Table 4. Overview of total costs during the 3-month follow-up for the complaint-directed mini-intervention (CDMI) and control groups[a].

Cost type	CDMI group (n=97)		Control group (n=140)	
	Mean cost (€)	95% CI	Mean cost (€)	95% CI
Intervention costs	4	—	0	—
Health care costs				
General practitioner visits	111	—	89	—
General practitioner support	8	—	5	—
Social worker	10	—	21	—
Psychologist	84	—	84	—
Physiotherapist	52	—	55	—
Psychiatry	91	—	75	—
Other visits[b]	40	—	68	—
Medication	151	—	224	—
Total health care costs[c]	519	428-611	600	458-741
Patient and family costs				
Home care	3	—	39	—
Special home care[b]	19	—	1	—
Informal care	100	—	83	—
Total patient and family costs[c]	114	77-151	126	38-214
Productivity losses				
Absenteeism paid work	943	—	911	—
Presenteeism paid work	560	—	625	—
Total production losses[c]	1461	1087-1835	1504	984-2024
Total costs after 3 months[c]	2094	1692-2496	2230	1679-2979

[a] Indexed for the year 2016.

[b] Includes alternative healing and self-support groups.

[c] Totals and subtotals based on multiple imputation estimates (CDMI group: n=165; control group: n=164).

Sensitivity Analyses

Limiting the analysis to patients with a paid job resulted in a dominant ICER at 3 months' follow-up (ie, lower costs and a higher rate of treatment response). When excluding baseline corrections, results were similar to the base case analyses (ie, dominant ICER at 3 months' follow-up). Including opportunity costs for patients also did not affect the ICER at 3 months. We found a dominant ICER when taking the health care perspective at 3 months' follow-up.

Looking at the costs per QALY gained, excluding patients without a paid job or an analysis without baseline correction resulted in a dominant ICUR for CDMI compared with the control group at 3 months' follow-up (ie, lower costs and increase in QALYs). Including opportunity costs for patients resulted in a dominant ICUR. From a health care perspective, correspondingly, we found a dominant ICUR at 3 months' follow-up.

Figure 1. Cost-effectiveness plane (left) and cost-effectiveness acceptability curve (right) of rates of responders to treatment at 3-month follow-up (costs per extra responder). LL: lower limit of the 95% CI; PE: mean incremental cost-effectiveness ratio (ICER); reps: ICER replication; UL: upper limit of the 95% CI.

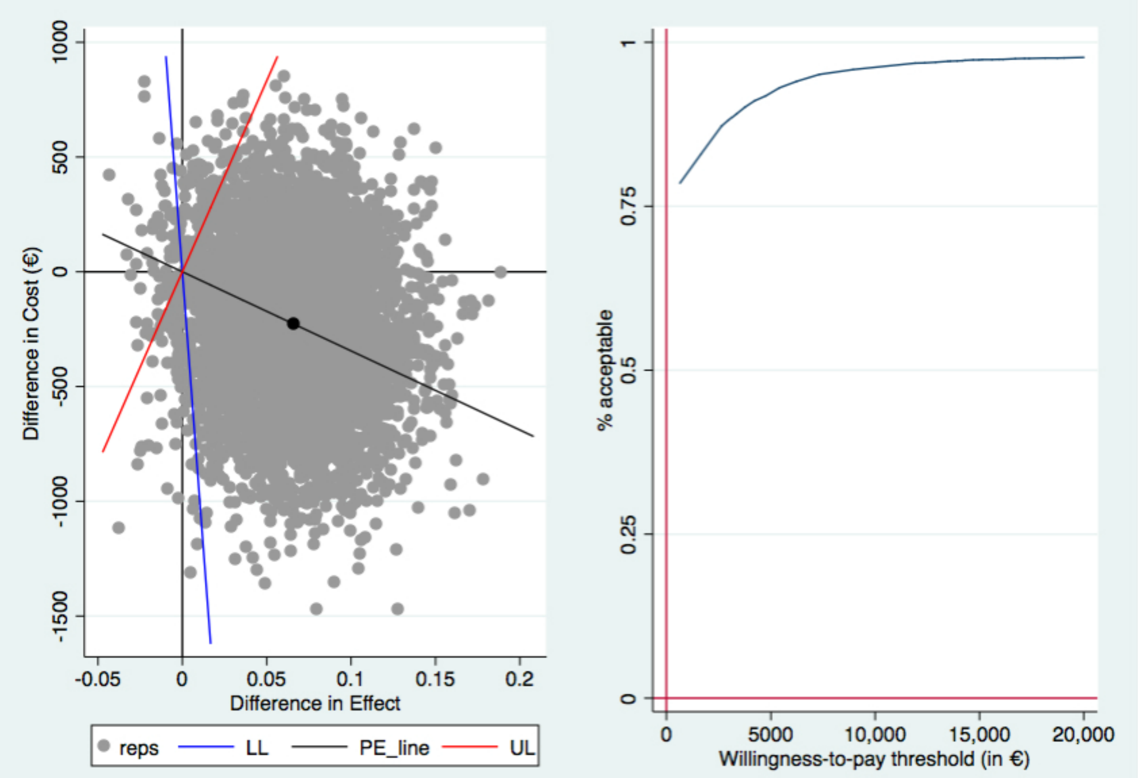


Figure 2. Cost-effectiveness plane (left) and cost-effectiveness acceptability curve (right) of quality-adjusted life-year (QALY) gain (costs per QALY gained) after 3 months. LL: lower limit of the 95% CI; PE: mean incremental cost-utility ratio (ICUR); reps: ICUR replication; UL: upper limit of the 95% CI.

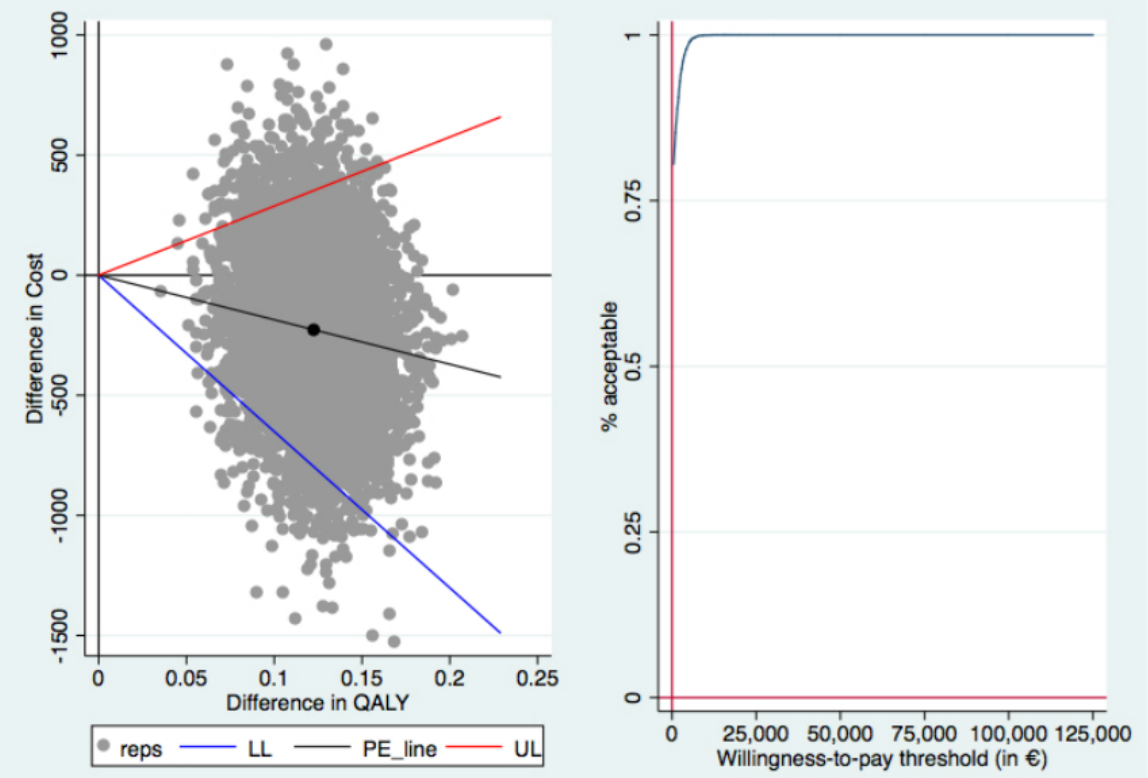


Table 5. Cost-effectiveness analyses and sensitivity analyses for rate of responders to treatment and quality-adjusted life-years (QALYs) at 3-month follow-up.

Analysis	Incremental cost (€)	Incremental effect	Mean ICER[a]	Distribution of 5000 bootstrap simulated ICERs			
				NE[b]	SE[c] (dominant)	SW[d]	NW[e] (inferior)
Cost effectiveness, rate of responders to treatment							
Main analysis	–225	0.07	Dominant	25.2	72.4	1.6	0.8
Only including participants with paid job	–309	0.12	Dominant	25.0	74.6	0.3	0.1
Analysis without baseline adjustments	–131	0.07	Dominant	33.7	63.9	1.4	1.0
Health care perspective	–78	0.07	Dominant	17.1	80.5	1.9	0.5
Including opportunity costs for participants	–225	0.07	Dominant	25.2	72.4	1.6	0.8
Regression-based imputation	–127	0.07	Dominant	34.4	63.3	1.3	1.0
Last observation carried forward	–784	0.05	Dominant	4.5	90.5	4.7	0.3
Cost utility, QALYs							
Main analysis	–225	0.12	Dominant	24.7	75.3	0	0
Only including participants with paid job	–312	0.15	Dominant	24.9	75.1	0	0
Analysis without baseline adjustments	–131	0.12	Dominant	34.7	65.3	0	0
Health care perspective	–85	0.12	Dominant	16.1	83.9	0	0
Including opportunity costs for participants	–228	0.12	Dominant	25.7	74.3	0	0
Regression-based imputation	–130	0.12	Dominant	35.1	64.9	0	0
Last observation carried forward	–785	0.06	Dominant	4.8	95.0	0.2	0.1

[a] ICER: incremental cost-effectiveness ratio.

[b] NE: northeast quadrant (the intervention was more effective and more costly than usual care).

[c] SE: southeast quadrant (the intervention was more effective and less costly than usual care).

[d] SW: southwest quadrant (the intervention was less effective and less costly than usual care).

[e] NW: northwest quadrant (the intervention was less effective and more costly than usual care).

Using covariates significantly associated with dropout at 3 months (ie, baseline GAD-7 score and age) and covariates significantly associated with the primary outcome of rate of responders to treatment on the IDS-R (ie, condition, paid work, baseline JSEQ score, and baseline PSS score) to impute missing values resulted in a dominant ICER and ICUR for CDMI compared with the control group at 3 months' follow-up. Using last observation carried forward to impute missing values also resulted in a dominant ICER and ICUR for CDMI compared with the control group at 3 months' follow-up. All in all, the sensitivity analyses attested to the robustness of the findings of the main analysis (see Table 5).

Discussion

Principal Findings

This study examined the cost effectiveness of Web-based CDMIs in adult patients with mild to moderate depressive symptoms in comparison with a wait-listed control group with unrestricted access to usual care in the Netherlands. The study had a follow-up at 3 months and in the experimental arm of the trial an extended follow-up at 6 months to see whether effects were sustained over time. The CDMI consisted of 3 different interventions (Sleep better, Stress less, and Worry less). Patients in the waiting-list control group were given access to the CDMIs after 3 months' follow-up. When looking at the rate of responders to treatment (defined as a 50% reduction in IDS-SR depressive symptoms), we found a dominant ICER at 3 months, implying that the CDMIs provided lower costs for better rates of responders to treatment. Looking at costs per QALY gained, we found a dominant ICER at 3 months, implying lower costs and increased QALYs. For both outcomes, results were sustainable over 6 months, particularly given the steady increase in the rate of responders to treatment in the CDMI group at 6 months and the increase in this responder rate in the control group at 6 months (after giving them access to the intervention at 3 months. Sensitivity analyses showed that results were robust to different assumptions, perspectives, or the way missing data was handled. Hence, this study demonstrated that it is possible to use an easily accessible and economically affordable intervention to improve participants' health status in a cost-effective manner.

Evidence in Context

A recent systematic review looking specifically at internet- and mobile-based interventions targeting depression highlighted the potential of those interventions to be cost effective, with cost-effectiveness ratios similar to those reported for face-to-face psychotherapy and antidepressant drug treatment [35]. In this review, of the 14 e-interventions, 6 were deemed cost effective, 5 were not cost effective, and 2 were undecided.

A systematic review looking at economic evaluations of internet interventions for mental health concluded that guided internet interventions for, among others, depression and anxiety demonstrated higher probabilities of being cost effective [36]. However, the evidence for unguided internet interventions for depression was less convincing.

A study examining the cost effectiveness of a Web-based self-help intervention aimed at enhancing well-being by fostering positive emotions and stimulating positive functioning demonstrated reduced depressive symptoms, although at higher costs, leading to unfavorable cost-effectiveness ratios [37]. The authors emphasized the importance of adherence to maintain long-term effects and possibly increase the cost effectiveness.

In terms of feasibility, Griffiths and Christensen [38] evaluated 2 community-based internet programs in the treatment of depression and concluded that these intervention programs could be delivered effectively over the internet. Furthermore, they emphasize the importance of using the internet as a more accessible alternative than face-to-face mental health services, especially in rural areas [38]. CDMIs are well suited to prevent health status deterioration at an early stage, especially in the current climate of increasing emphasis on self-reliance and self-management.

Strengths and Limitations

This study was not without limitations. First, the follow-up of the trial was 3 months after baseline, which did not provide hard evidence for the longer-term effects. However, in the extended follow-up of the experimental arm, we could see that the effects (rate of responders to treatment and QALY gains) were maintained at 6 months. In addition, at the 6-month follow-up, the costs were relatively lower than those at 3 months (for both groups), suggesting that health care resource use may have decreased. Second, dropout rates were relatively large, which made imputation of missing values necessary. Although high dropout rates are a problem often encountered in electronic health (eHealth) trials [39], one should always carefully consider imputation techniques, especially in the context of substantial dropout. However, different imputation techniques led to comparable results, attesting to the robustness of the main analysis. Third, given the nature of the intervention, participants could not be blinded. This may have biased participants in one way or another (eg, placebo effect). Fourth, because we recruited participants mainly by means of internet-based recruitment avenues, it is possible that we missed some potential participants. For example, we may have missed participants who are less likely to engage in social media. Fifth, although participants were free to find additional care themselves, we did not provide the control condition with any intervention. This may have caused an overestimation of the effects of the CDMIs. Sixth, given the focus of each CDMI, some of the content differed between the 3 CDMIs, although there was also overlap in content. The CDMIs can be seen as an overall intervention approach that aims to target depressive complaints, but one in which participants are able to choose the CDMI they want to use based on their personal needs, and they do not have to use CDMIs that are not relevant to their situation. As a result, each CDMI may target depressive complaints differently. This would be an interesting avenue to explore in future to gain insight into the mechanisms of change. Seventh, the generalizability of the findings to men and those with other educational levels remains to be determined, as mainly highly educated female participants were included in the trial. However, this selected group of participants may well reflect the composition of the target group that will be reached after implementation of the CDMIs.

Conclusion

This study demonstrated that the brief and low-threshold Web-based, unguided, self-help CDMIs have the potential to be a cost-effective addition to usual care for adults with mild to moderate depressive symptoms. The CDMIs were shown to improve health status, while at the same time reducing health care costs of participants, and hence dominating the care-as-usual control condition. As intervention costs were relatively low, and the internet is nowadays readily available in the Western world, we believe the CDMIs can be easily implemented on a large scale. Future research should aim at increasing the reach of the intervention and determining whether the intervention is indeed more likely to reach people with a low socioeconomic status. Related to this matter, integration of the CDMIs into primary care may be a useful next step, as this would allow the CDMIs to be offered with some guidance from the general practice nurse, possibly boosting effectiveness and adherence. Regarding adherence, future research may also focus on the impact of reminder systems incorporated into the CDMIs.

Acknowledgments

This study was made possible with financial support from the Netherlands Ministry of Health, Welfare and Sport and the Trimbos Institute. We would like to thank all the study participants. Our thanks is also extended to Hester Duivis, Marijke Ruiter, and Barbara Conijn for their efforts in the development of the interventions and the study design.

Conflicts of Interest

The interventions that were used in this study were developed by the Trimbos Institute, Netherlands Institute of Mental Health and Addiction, a nonprofit organization, where S Lokman, S Leone, FS, and BFMW were employed at the time of the study. All authors declare that they have no other conflicts of interest.

Multimedia Appendix 1

Detailed insights of complaint-directed mini-intervention (CMDI) including screenshots.

[PDF File (Adobe PDF File), 690KB - jmir_v20i10e10455_app1.pdf]

Multimedia Appendix 2

CONSORT‐EHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 647KB - jmir_v20i10e10455_app2.pdf]

References

1. World Health Organization. Depression. 2018 Mar 22. URL: http://www.who.int/en/news-room/fact-sheets/detail/depression [accessed 2018-09-11] [WebCite Cache ID 72LdMtX4s]
2. De Graaf R, Ten Have M, van Dorsselaer S. De psychische gezondheid van de Nederlandse bevolking. Nemesis-2: Opzet en eerste resultaten. Utrecht, Netherlands: Trimbos Institute; 2010.
3. Cuijpers P, Smit F. Subthreshold depression as a risk indicator for major depressive disorder: a systematic review of prospective studies. Acta Psychiatr Scand 2004 May;109(5):325-331. [doi: 10.1111/j.1600-0447.2004.00301.x] [Medline: 15049768]
4. Lokkerbol J, Adema D, de Graaf R, ten Have M, Cuijpers P, Beekman A, et al. Non-fatal burden of disease due to mental disorders in the Netherlands. Soc Psychiatry Psychiatr Epidemiol 2013 Oct;48(10):1591-1599. [doi: 10.1007/s00127-013-0660-8] [Medline: 23397319]
5. Smit F, Cuijpers P, Oostenbrink J, Batelaan N, de Graaf R, Beekman A. Costs of nine common mental disorders: implications for curative and preventive psychiatry. J Ment Health Policy Econ 2006 Dec;9(4):193-200. [Medline: 17200596]
6. Volksgezondheidenzorg. Stemmingsstoornissen, Kosten. 2017 Oct 10. URL: https://www.volksgezondheidenzorg.info/onderwerp/stemmingsstoornissen/kosten/kosten%20-%20node-kosten-van-zorg-voor-depressie [accessed 2018-09-11] [WebCite Cache ID 72LdjS4et]
7. Greenberg PE, Fournier A, Sisitsky T, Pike CT, Kessler RC. The economic burden of adults with major depressive disorder in the United States (2005 and 2010). J Clin Psychiatry 2015 Feb;76(2):155-162 [FREE Full text] [doi: 10.4088/JCP.14m09298] [Medline: 25742202]
8. Schrier AC, de Wit MAS, Rijmen F, Tuinebreijer WC, Verhoeff AP, Kupka RW, et al. Similarity in depressive symptom profile in a population-based study of migrants in the Netherlands. Soc Psychiatry Psychiatr Epidemiol 2010 Oct;45(10):941-951. [doi: 10.1007/s00127-009-0135-0] [Medline: 19763363]
9. Smit F, Majo C, Boon B, Ruiter M, Abspoel M. Klachtgerichte mini-interventisn: Een nieuwe aanpak van depressiepreventie voor groepen met een lagere sociaal-economische status. Utrecht, Netherlands: Trimbos Institute; 2011.
10. Borsboom D, Cramer AOJ. Network analysis: an integrative approach to the structure of psychopathology. Annu Rev Clin Psychol 2013;9:91-121. [doi: 10.1146/annurev-clinpsy-050212-185608] [Medline: 23537483]
11. McEvoy PM, Salmon K, Hyett MP, Jose PE, Gutenbrunner C, Bryson K, et al. Repetitive negative thinking as a transdiagnostic predictor of depression and anxiety symptoms in adolescents. Assessment 2017 Feb 01:1073191117693923. [doi: 10.1177/1073191117693923] [Medline: 29214855]
12. Newby JM, McKinnon A, Kuyken W, Gilbody S, Dalgleish T. Systematic review and meta-analysis of transdiagnostic psychological treatments for anxiety and depressive disorders in adulthood. Clin Psychol Rev 2015 Aug;40:91-110 [FREE Full text] [doi: 10.1016/j.cpr.2015.06.002] [Medline: 26094079]
13. Baglioni C, Spiegelhalder K, Nissen C, Riemann D. Clinical implications of the causal relationship between insomnia and depression: how individually tailored treatment of sleeping difficulties could prevent the onset of depression. EPMA J 2011 Sep;2(3):287-293 [FREE Full text] [doi: 10.1007/s13167-011-0079-9] [Medline: 23199164]
14. Wiegner L, Hange D, Björkelund C, Ahlborg G. Prevalence of perceived stress and associations to symptoms of exhaustion, depression and anxiety in a working age population seeking primary care--an observational study. BMC Fam Pract 2015;16:38 [FREE Full text] [doi: 10.1186/s12875-015-0252-7] [Medline: 25880219]
15. Topper M, Emmelkamp PM, Ehring T. Improving prevention of depression and anxiety disorders: repetitive negative thinking as a promising target. Appl Prev Psychol 2010 Jun;14(1-4):57-71. [doi: 10.1016/j.appsy.2012.03.001]
16. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. Methods for the Economic Evaluation of Health Care Programmes. Oxford, UK: Oxford University Press; 2015.
17. Zorginstituut Nederland. Economische evaluatie. 2016 Nov 15. URL: https://www.zorginstituutnederland.nl/over-ons/werkwijzen-en-procedures/adviseren-over-en-verduidelijken-van-het-basispakket-aan-zorg/beoordeling-van-geneesmiddelen/richtlijnen-voor-economische-evaluatie [accessed 2018-09-11] [WebCite Cache ID 72Lebdxv8]
18. Williams I, McIver S, Moore D, Bryan S. The use of economic evaluations in NHS decision-making: a review and empirical investigation. Health Technol Assess 2008 Apr;12(7):iii, ix-iiix, 1 [FREE Full text] [Medline: 18373906]
19. McCrone P, Knapp M, Proudfoot J, Ryden C, Cavanagh K, Shapiro DA, et al. Cost-effectiveness of computerised cognitive-behavioural therapy for anxiety and depression in primary care: randomised controlled trial. Br J Psychiatry 2004 Jul;185:55-62 [FREE Full text] [Medline: 15231556]

20. Warmerdam L, Smit F, van Straten A, Riper H, Cuijpers P. Cost-utility and cost-effectiveness of internet-based treatment for adults with depressive symptoms: randomized trial. J Med Internet Res 2010;12(5):e53 [FREE Full text] [doi: 10.2196/jmir.1436] [Medline: 21169166]
21. Lokman S, Leone SS, Sommers-Spijkerman M, van der Poel A, Smit F, Boon B. Complaint-directed mini-interventions for depressive complaints: a randomized controlled trial of unguided web-based self-help interventions. J Med Internet Res 2017 Jan 04;19(1):e4 [FREE Full text] [doi: 10.2196/jmir.6581] [Medline: 28052840]
22. Rush AJ, Gullion CM, Basco MR, Jarrett RB, Trivedi MH. The Inventory of Depressive Symptomatology (IDS): psychometric properties. Psychol Med 1996 May;26(3):477-486. [Medline: 8733206]
23. Sanderson K, Andrews G, Corry J, Lapsley H. Using the effect size to model change in preference values from descriptive health status. Qual Life Res 2004 Sep;13(7):1255-1264. [doi: 10.1023/B:QURE.0000037482.92757.82] [Medline: 15473504]
24. Zorginstituut Nederland. Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg. Diemen, Netherlands: Zorginstituut Nederland; 2015.
25. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, CHEERS Task Force. Consolidated Health Economic Evaluation Reporting Standards (CHEERS) statement. Cost Eff Resour Alloc 2013 Mar 25;11(1):6 [FREE Full text] [doi: 10.1186/1478-7547-11-6] [Medline: 23531194]
26. Hakkaart-van Roijen L, Straten AV, Tiemens B, Donker MCH. Trimbos/iMTA Questionnaire for Costs Associated With Psychiatric Illness (TIC-P). Rotterdam, Netherlands: Institute for Medical Technology Assessment; 2002.
27. Dutch Healthcare Institute. Medicijn Kosten. 2017. URL: https://www.medicijnkosten.nl [accessed 2018-09-11] [WebCite Cache ID 72LfROgeO]
28. Dutch Healthcare Institute. Farmacotherapeutisch Kompas. 2017. URL: https://www.farmacotherapeutischkompas.nl [accessed 2018-03-20] [WebCite Cache ID 6y3wlrrZ1]
29. Grittner U, Gmel G, Ripatti S, Bloomfield K, Wicki M. Missing value imputation in longitudinal measures of alcohol consumption. Int J Methods Psychiatr Res 2011 Mar;20(1):50-61 [FREE Full text] [doi: 10.1002/mpr.330] [Medline: 21556290]
30. Horton NJ, Lipsitz SR. Multiple imputation in practice: comparison of software packages for regression models with missing variables. Am Stat 2001 Aug;55(3):244-254. [doi: 10.1198/000313001317098266]
31. Obenchain RL, Melfi CA, Croghan TW, Buesching DP. Bootstrap analyses of cost effectiveness in antidepressant pharmacotherapy. Pharmacoeconomics 1997 May;11(5):464-472. [Medline: 10168034]
32. Council for Public Health and Health Care. [Sensible and sustainable care]. Zoetermeer, Netherlands: The Council; 2006.
33. Birch S, Gafni A. On being NICE in the UK: guidelines for technology appraisal for the NHS in England and Wales. Health Econ 2002 Apr;11(3):185-191. [Medline: 11921316]
34. Demirtas H. Simulation driven inferences for multiply imputed longitudinal datasets. Stat Neerlandica 2004 Nov;58(4):466-482. [doi: 10.1111/j.1467-9574.2004.00271.x]
35. Paganini S, Teigelkötter W, Buntrock C, Baumeister H. Economic evaluations of internet- and mobile-based interventions for the treatment and prevention of depression: a systematic review. J Affect Disord 2018 Jan 01;225:733-755. [doi: 10.1016/j.jad.2017.07.018] [Medline: 28922737]
36. Donker T, Blankers M, Hedman E, Ljótsson B, Petrie K, Christensen H. Economic evaluations of Internet interventions for mental health: a systematic review. Psychol Med 2015 Dec;45(16):3357-3376. [doi: 10.1017/S0033291715001427] [Medline: 26235445]
37. Bolier L, Majo C, Smit F, Westerhof G, Haverman M, Walburg J, et al. Cost-effectiveness of online positive psychology: randomized controlled trial. J Posit Psychol 2014;9(5):460-471.
38. Griffiths KM, Christensen H. Internet-based mental health programs: a powerful tool in the rural medical kit. Aust J Rural Health 2007 Apr;15(2):81-87. [doi: 10.1111/j.1440-1584.2007.00859.x] [Medline: 17441815]
39. 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]

Abbreviations

CDMI: complaint-directed mini-intervention
eHealth: electronic health
GAD-7: Generalized Anxiety Disorder 7-item
ICER: incremental cost-effectiveness ratio
ICUR: incremental cost-utility ratio
IDS-SR: Inventory of Depressive Symptomatology Self-Report
JSEQ: Jenkins Sleep Evaluation Questionnaire
PSS: Perceived Stress Scale
QALY: quality-adjusted life-years
SURE: seemingly unrelated regression equations

Edited by T Cosco; submitted 22.03.18; peer-reviewed by A Parks, M Krausz; comments to author 11.04.18; revised version received 30.05.18; accepted 22.07.18; published 01.10.18.

Please cite as:
Wijnen BFM, Lokman S, Leone S, Evers SMAA, Smit F
Complaint-Directed Mini-Interventions for Depressive Symptoms: A Health Economic Evaluation of Unguided Web-Based Self-Help Interventions Based on a Randomized Controlled Trial
J Med Internet Res 2018;20(10):e10455
URL: https://www.jmir.org/2018/10/e10455/
doi:10.2196/10455
PMID:30274958

Original Paper

Tailored, Therapist-Guided Internet-Based Cognitive Behavioral Therapy Compared to Care as Usual for Patients With Rheumatoid Arthritis: Economic Evaluation of a Randomized Controlled Trial

Maaike Ferwerda[1,2], PhD; Sylvia van Beugen[1,2], PhD; Henriët van Middendorp[1,2], PhD; Henk Visser[3], MD, PhD; Harald Vonkeman[4,5], MD, PhD; Marjonne Creemers[6], MD, PhD; Piet van Riel[7], MD, PhD; Wietske Kievit[8], PhD; Andrea Evers[1,2], PhD

[1]Health, Medical and Neuropsychology Department, Institute of Psychology, Leiden University, Leiden, Netherlands
[2]Medical Psychology Department, Radboud University Medical Center, Nijmegen, Netherlands
[3]Department of Rheumatology, Rijnstate Hospital, Arnhem, Netherlands
[4]University of Twente, Enschede, Netherlands
[5]Arthritis Center Twente, Department of Rheumatology and Clinical Immunology, Medisch Spectrum Twente, Enschede, Netherlands
[6]Department of Rheumatology, Jeroen Bosch Hospital, Den Bosch, Netherlands
[7]Scientific Institute for Quality of Healthcare, Radboud University Medical Center, Nijmegen, Netherlands
[8]Department for Health Evidence, Radboud University Medical Center, Nijmegen, Netherlands

Corresponding Author:
Maaike Ferwerda, PhD
Health, Medical and Neuropsychology Department
Institute of Psychology
Leiden University
PO Box 9555
Leiden, 2300 RB
Netherlands
Phone: 31 71 5273627
Fax: 31 71 5273619
Email: m.ferwerda@fsw.leidenuniv.nl

Abstract

Background: Internet-based cognitive behavioral therapy can aid patients with rheumatoid arthritis with elevated levels of distress to enhance their quality of life. However, implementation is currently lacking and there is little evidence available on the (cost-) effectiveness of different treatment strategies.

Objective: Cost-benefit ratios are necessary for informing stakeholders and motivating them to implement effective treatment strategies for improving health-related quality of life (HRQoL) of patients with rheumatoid arthritis. A cost-effectiveness study from a societal perspective was conducted alongside a randomized controlled trial on a tailored, therapist-guided internet-based cognitive behavioral therapy (ICBT) intervention for patients with rheumatoid arthritis with elevated levels of distress as an addition to care as usual (CAU).

Methods: Data were collected at baseline or preintervention, 6 months or postintervention, and every 3 months thereafter during the 1-year follow-up. Effects were measured in terms of quality-adjusted life years (QALYs) and costs from a societal perspective, including health care sector costs (health care use, medication, and intervention costs), patient travel costs for health care use, and costs associated with loss of labor.

Results: The intervention improved the quality of life compared with only CAU (Δ QALYs=0.059), but at a higher cost (Δ=€4211). However, this increased cost substantially reduced when medication costs were left out of the equation (Δ=€1863). Of all, 93% (930/1000) of the simulated incremental cost-effectiveness ratios were in the north-east quadrant, indicating a high probability that the intervention was effective in improving HRQoL, but at a greater monetary cost for society compared with only CAU.

Conclusions: A tailored and guided ICBT intervention as an addition to CAU for patients with rheumatoid arthritis with elevated levels of distress was effective in improving quality of life. Consequently, implementation of ICBT into standard health care for patients with rheumatoid arthritis is recommended. However, further studies on cost reductions in this population are warranted.

Trial Registration: Nederlands Trial Register NTR2100; http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=2100 (Archived by WebCite at http://www.webcitation.org/724t9pvr2)

(J Med Internet Res 2018;20(10):e260) doi:10.2196/jmir.9997

KEYWORDS

cost-utility analysis; cognitive behavioral treatment; Internet-based therapy; rheumatoid arthritis

Introduction

The psychological impact of rheumatoid arthritis has become increasingly apparent, with patients reporting decreased health-related quality of life (HRQoL) as a result of physical factors, such as pain, and psychological factors, such as negative mood [1-3]. As these factors are associated with the disease trajectory, health care utilization, and workplace disability of patients [4-8], they often lead to significant societal health expenses [9-12].

Approximately one-third patients with rheumatoid arthritis experience a significantly reduced HRQoL [1,3], and cognitive behavioral therapy can aid in improving HRQoL [13-15]. In a recent randomized controlled trial (RCT), we demonstrated that a therapist-guided internet-based cognitive behavioral therapy intervention (ICBT) that was tailored to specific problems of individual patients with rheumatoid arthritis with elevated levels of distress led to improvements in their psychological functioning (eg, depressed mood) [16]. Findings of this study were in agreement with those of previous studies on face-to-face cognitive behavioral therapies [17], which our treatment protocol closely resembled. Furthermore, it also supported previous preliminary evidence suggesting that ICBT is as effective as face-to-face treatments for a range of somatic conditions and symptoms [18-20].

Several benefits of internet-based treatments, including increased flexibility in terms of time and place, may make these treatments feasible for widespread implementation [21,22]. Although evidence on cost-effectiveness of internet-based therapy is still scarce, preliminary results suggest that such interventions are a cost-effective method to improve mental health, specifically when guidance is provided by a psychological therapist [23]. One study examining a self-management intervention for patients with RA reported a reduction in general distress and pain, and improvement in self-efficacy, although no effects on health care utilization were seen [24], which is important for evaluation of costs. Furthermore, improved quality of life in patients with chronic somatic conditions has previously been associated with an improvement in medication adherence, self-efficacy, and positive health outcomes [10,12]. This could potentially reduce the cost of health care through, for example, greater adherence to medications and increased employability and work outcomes. However, the evidence on this is still scarce.

Specifically, no studies have examined the cost-effectiveness of ICBT in patients with rheumatoid arthritis thus far. This information is essential for allowing stakeholders to balance treatment choices and policy decisions. For example, a recent study summarized how rheumatologists balanced multiple aspects of a treatment choice, including efficacy, patient preferences, and costs [25].

This study describes a preplanned cost-effectiveness study conducted from a societal perspective on the use of a tailored, therapist-guided ICBT protocol as an addition to care as usual (CAU) for patients with rheumatoid arthritis with elevated levels of distress. This was conducted alongside RCT, the results of which have been reported elsewhere [16]. We predicted that ICBT would be a cost-effective intervention as addition to CAU. In particular, costs that are relevant to society, that is, health care sector costs such as medication costs, health care usage costs, and work-related costs, were examined as these potentially decrease with improving HRQoL.

Methods

Design

An economic evaluation of a tailored, therapist-guided ICBT protocol as an addition to CAU was conducted from a societal perspective alongside an RCT. Patients with elevated levels of distress were randomly selected to receive standard rheumatological care (as usually conducted in the Netherlands) only or in combination with ICBT. Further details of the RCT can be found in a previous study reporting the effects of ICBT on the psychological functioning, physical functioning, and impact of rheumatoid arthritis on daily life [16]. This study focuses only on aspects relevant to economic evaluation. All patients provided written informed consent for participation in the study. The regional medical ethical committee approved the study (NL24343.091.08), and it was registered with the Nederlands Trial Register (NTR2100).

Participants

Adult patients with a rheumatologist-certified diagnosis of rheumatoid arthritis [26] and receiving out-patient standard rheumatological care at 1 academic and 3 nonacademic hospitals were invited to participate in this study. Only patients with elevated levels of distress, as defined by high scores for negative mood (≥21 for negative mood on the Impact of Rheumatic Diseases on General Health and Lifestyle scale) [27] and/or anxiety (a score of ≥5 for anxiety on Impact of Rheumatic Diseases on General Health and Lifestyle scale) were included. The exclusion criteria were (1) insufficient command of the Dutch language, (2) severe physical or psychiatric comorbidity (ie, requiring acute or intensive medical attention; when this was not the case, patients indicated which condition impacted their HRQoL to a greater extent), (3) pregnancy, (4) currently receiving treatment from a cognitive behavioral therapist or comparable practitioner, and (5) no access to a computer and internet.

Care as Usual and Internet-Based Cognitive Behavioral Intervention

CAU for patients with rheumatoid arthritis, which was provided to the intervention and control groups, generally consists of shared care checkups provided every 3-6 months by a rheumatology nurse and the rheumatologist to monitor disease activity and treatment. Hospitals in the Netherlands follow the recommendations for rheumatological care provided by the Dutch Society for Rheumatology. In addition, physical and occupational therapy may be provided, depending on the patient and disease characteristics.

The intervention group received ICBT as an addition to CAU. The treatment was tailored to individual patient goals and characteristics and guided by a therapist. The treatment commenced with 1 or 2 face-to-face intake sessions comprising formulation of individual goals based on the main problems of the patient. Based on these goals, specific treatment modules embedded within the ICBT website were selected, and the therapist guided the choice of assignments within each of these modules based on the specific risk and resilience factors of the patient. Therapists and patients remained in contact weekly or biweekly (based on patient preferences) via a secured messaging service within the ICBT website. Treatment modules focused on coping with (1) pain and functional disability, (2) fatigue, (3) social functioning, and (4) negative mood. As the modules were tailored to individual requirements, treatment durations varied from 9 to 65 (mean 26.07, SD 12.22) weeks. All 6 therapists had a master's degree in psychology and 2 had additional postacademic training in cognitive behavioral therapy. Supervision was provided by a senior clinical psychologist with postacademic training in cognitive behavioral therapy. Patients received 1 telephonic session that lasted for 30 minutes, where a research assistant explained how the intervention website was set up. Further information on the ICBT intervention can be found in our previous study [16].

Data Collection and Outcome Measures

Data were collected at baseline; postintervention for the intervention group and 6 months after baseline for the control group; and at 3 (F1), 6 (F2), 9 (F3), and 12 months (F4) thereafter. All questionnaires were filled out in paper and pencil versions. All costs were calculated based on the 2015 Dutch price indices. The last observation carried forward was applied for missing data to account for biases introduced by nonresponse.

Effects: Quality-Adjusted Life Years

HRQoL was assessed using the Dutch version of the EuroQol-5dimensions-3levels (EQ-5D-3L) questionnaire [28]. EQ-5D-3L captures 5 dimensions of health, including mobility, self-care, usual activities, pain or discomfort, and anxiety or depression. Each dimension has 3 response options: no, some or moderate, and extreme problems. Utility scores were calculated using the Dutch tariff [28], with scores ranging from 0 (death) to 1 (perfect health). The trapezium rule was applied for calculating area under the curve for measuring quality-adjusted life years (QALYs).

Costs: Societal Perspective

Costs were calculated for 3 dimensions, including health care sector costs (comprising health care use, medication, and costs of the intervention under study), patient travel costs for health care use, and costs associated with loss of labor (absenteeism and presenteeism).

Health care use was assessed using the Trimbos or Institute for Medical Technology Assessment questionnaire for Costs associated with Psychiatric Illness (TiC-P) [29], which was adjusted for health care use by patients with rheumatoid arthritis. The questionnaire included patient appointments with rheumatologists, specialized rheumatology nurses, occupational therapists, physical therapists, podiatrists, and hydrotherapists; admissions to daycare; and inpatient treatment at hospitals or rehabilitation centers. Furthermore, TiC-P also assessed care provided by a general practitioner and occupational health doctor, psychological or psychosocial care (eg, care provided by a psychologist, psychiatrist, or social worker), and care provided by alternative medicine practitioners. Costs were calculated by multiplying health care use with estimates of unit prices, as provided by the TiC-P [29] and the Dutch manual for cost analyses in health care [30].

All medications related to rheumatoid arthritis were taken into account, including pain medication (eg, nonsteroidal antiinflammatory drugs), corticosteroids, disease modifying antirheumatic drugs, and biologics. Furthermore, medications related to psychological symptoms, such as depression and anxiety, as well as those related to sleep disorders were also taken into consideration. Costs of these medications were calculated by multiplying dosages for each type of medication with their costs as per the Dutch national tariff list. The medication history was self-reported by the patients.

In accordance with the Dutch guidelines for cost analyses in health care [30], costs for the tailored, therapist-guided ICBT protocol were calculated using (1) the actual costs of development of the intervention by the ICT-company, which included updating and security costs, (2) salary of the therapists (as per rates for basic psychologists and those with a postdoc training, where appropriate) based on the amount of time they spent on treatment for each patient (including face-to-face intake sessions, internet-based communication, and additional telephone calls), (3) salary costs for the research assistant conducting the telephone session to explain the intervention website, and (4) patient traveling expenses for the face-to-face intake sessions. An amortization period of 5 years was assumed. Costs per patient were calculated based on prevalence rates of rheumatoid arthritis provided by the Netherlands Institute for Health Services Research [31]. An assumption was made that 30% of the rheumatoid arthritis population is eligible for this intervention because of elevated levels of distress, as observed in the RCT [16] and a previous trial targeting the same population [17]. Of this population size, a population reach of 10% was assumed.

Patient travel costs were calculated by multiplying the Dutch standard for average travel distances from home to several health care services (for example, hospital, general practitioner, and physical therapist), as per the Dutch manual for cost analysis in health care [30], with a price of €0.19 per kilometer.

Loss of productivity costs were calculated for 1 year using the friction costs method including presenteeism and absenteeism, based on self-reported data on loss of productivity collected via the PROductivity and DIsease Questionnaire (PRODISQ) [32]. The friction period was calculated to a maximum of 12 weeks [30]. An additional period of 4 weeks was taken into consideration to allow management to fill the vacancy. Loss of productivity costs was calculated by multiplying overall average costs of productivity loss per hour (€34.90) [30] with the number of hours that a patient was absent from work or was unable to perform optimally at work because of rheumatoid arthritis.

Statistical Analysis

Differences in baseline sociodemographic (eg, age and gender), disease-related (eg, disease severity), and economic (eg, paid labor, health care costs, medication costs, and HRQoL) characteristics between the intervention and control groups were assessed using independent sample *t* tests or chi-square tests as appropriate.

The costs per QALY gained were assessed using an incremental cost-utility ratio (ICUR), calculated by dividing the difference in costs by the difference in QALYs. Bootstrapping (1000 replications) was used to nonparametrically determine 95% CI. Results of the bootstrap were presented and analyzed by means of a cost-effectiveness plane and willingness-to-pay curve. The Dutch Council for Public and Health Care (RVZ) recommends that the threshold of the ICUR in relation to the acceptability of the treatment has to depend on the severity of the disease with a maximum ICUR of €80,000/QALY [33]. In accordance, the probability that this intervention remains within this threshold for willingness to pay is reported. As the intervention was not primarily aimed at reducing medication costs and a substantial proportion of patients with rheumatoid arthritis used expensive biologic agents that strongly influenced cost estimations, a secondary analysis was performed where costs for medication were excluded.

Results

Patient Demographics

Patient demographics have been shown in Table 1 and Table 2. This study included 133 patients, of which 62 patients were in the ICBT group and 71 were in the CAU group. Measurements at all time-points were provided by 27% (17/62) patients allocated to the ICBT group and 42% (30/71) patients allocated to the CAU group. No baseline differences in demographics, disease-related characteristics, and cost- or effect-related variables were observed. The patient sample, which included more female than male patients (85 female and 48 male) and had a mean age of 56.35 (SD 10.00; range: 26-81) years, was a representative of patients with rheumatoid arthritis.

Between-Group Differences in Effects: Quality-Adjusted Life Years

An overview of EQ-5D-3L utility scores for the intervention and control groups has been shown in Table 3, based on a 1-year follow-up period, indexed to the year 2015, for primary and secondary analyses. Although HRQoL was found to be similar for the intervention and control groups at baseline (*P*=.16), the former exhibited a higher QALY score (mean QALY=0.86, 2.5-97.5 percentile=0.82-0.89) than the latter (mean QALY=0.80, 2.5-97.5 percentile=0.76-0.83) during the 1-year follow-up period.

Between-Group Differences in Costs: Societal Perspective

The intervention and control groups did not exhibit any differences with regard to all costs during the 1-year follow-up period (Table 4), based on a 1-year follow-up period, indexed to the year 2015. Total costs for the intervention amounted to €419 per patient.

Table 1. Baseline patient characteristics: continuous variables.

Group	Characteristics			
	EQ-5D-3L[a]		Disease activity (RADAI[b])	
	Mean (SD)	Range	Mean (SD)	Range
CAU[c] (N=71)	0.69 (0.23)	−0.11 to 1.0	3.84 (1.75)	0.40 to 7.27
ICBT[d] + CAU (N=62)	0.74 (0.19)	0.09 to 1.0	3.31 (1.99)	0.20 to 7.95
Total group (N=133)	0.71 (0.21)	−0.11 to 1.0	3.59 (1.88)	0.20 to 7.95

[a] EQ-5D-3L: EuroQol-5dimensions-3 levels. Outcome analysis; group differences at baseline were analyzed using independent samples *t* tests as appropriate. *P*=.16.

[b] RADAI: Rheumatoid Arthritis Disease Activity Index. Group differences at baseline were analysed using independent samples *t* tests as appropriate. *P*=.11.

[c] CAU: care as usual.

[d] ICBT: internet-based cognitive behavioral therapy.

Table 2. Baseline patient characteristics: dichotomous variables.

Characteristics	CAU[a] (N=71)			ICBT[b] + CAU (N=62)			Total group (N=133)			*P* value[c]
	Yes	No	Missing	Yes	No	Missing	Yes	No	Missing	
Medical comorbidity	36	32	3	30	30	2	66	62	5	.74
Psychological comorbidity	5	63	3	2	58	2	7	121	5	.32
Employed	28	40	3	28	33	1	56	73	4	.59
Medication use										
Painkillers	8	56	7	10	46	6	18	102	13	.41
NSAIDS[d]	29	35	7	23	33	6	52	68	13	.64
DMARDS[e]	52	12	7	46	10	6	98	22	13	.90
Corticoids	13	51	7	12	44	6	25	95	13	.88
Biologicals	21	43	7	24	32	6	45	75	13	.26
Mental Health Medication	6	58	7	4	52	6	10	110	13	.66

[a] CAU: care as usual.

[b] ICBT: internet-based cognitive behavioral therapy.

[c] Outcome analysis; group differences at baseline were analyzed using chi-square analysis or independent samples *t* tests as appropriate.

[d] NSAID: Nonsteroidal antiinflammatory drugs.

[e] DMARD: disease modifying antirheumatic drugs.

Table 3. Quality-adjusted life years and costs for the care as usual and internet-based cognitive behavioral therapy plus care as usual groups.

Analysis	CAU[a]		ICBT[b] + CAU		Δ QALY[c]	Δ Costs
	QALY	Costs	QALY	Costs		
Primary analysis						
Average	0.80	€11,542	0.86	€15,754	0.059	€4211
2.5 percentile	0.76	€11,830	0.82	€8671	0.007	−€636
97.5 percentile	0.83	€20,134	0.89	€14,599	0.090	€9481
Secondary analysis, excluding medication costs						
Average	0.80	€2846	0.86	€4774	0.0590	€1863
2.5 percentile	0.76	€1743	0.82	€2541	0.007	−€714
97.5 percentile	0.83	€4243	0.89	€7777	0.090	€5428

[a] CAU: care as usual.

[b] ICBT: internet-based cognitive behavioral therapy.

[c] QALY: quality-adjusted life years.

Table 4. Mean costs for the care as usual and the internet-based cognitive behavioral therapy plus care as usual groups.

Cost category	CAU[a], mean (SD)	ICBT[b] + CAU, mean (SD)	Difference between groups *P* value
Health care use	€2548 (3659)	€3252 (8477)	.52
Medication use	€8682 (12,469)	€10,901 (13,257)	.32
Patient travel costs	€109 (135)	€151 (160)	.93
Absenteeism	€363 (1258)	€1309 (9106)	.89
Presenteeism	€1800 (5853)	€2239 (7133)	.91
ICBT intervention	N/A[c]	€419	N/A

[a] CAU: care as usual.

[b] ICBT: internet-based cognitive behavioral therapy.

[c] N/A: not applicable.

Results of the cost-utility analysis have been presented in Table 3. Incremental effectiveness resulted in an effect of 0.059, whereas incremental costs amounted to €4211.44 (2.5-97.5 percentile=−€636-9481). Therefore, the incremental cost-effectiveness amounted to an investment of €71,424.82 costs per QALY gained. The cost-effectiveness plane containing a scatterplot of simulated ICURs has been shown in Figure 1. Majority (93%, 930/1000) of simulated ICURs were in the north-east quadrant, indicating a high probability that the intervention was effective in improving HRQoL, but at a greater cost for society compared with CAU. A total of 6% (60/1000) of ICURs were in the south-east quadrant, suggesting greater HRQoL effects at lower costs to society, and 1% (10/1000) of ICURs were in the north-west quadrant, suggesting lower HRQoL effects at higher costs to society (Figure 1). At a willingness to pay of €80,000, the intervention had a 57% chance of being cost-effective (Figure 2) for patients with rheumatoid arthritis with elevated levels of distress.

Upon repeating the analysis without taking the medication costs into account, incremental costs were seen to reduce to €1862.72 (2.5-97.5 percentile=−€714-€5428). The scatterplot of the simulated ICURs remained approximately the same (Figure 3), excluding costs of medication; however, the cost-effectiveness acceptability curve exhibited an 87% chance of being cost-effective at a willingness to pay of €80,000 (Figure 4) for patients with rheumatoid arthritis with elevated levels of distress, without taking medication costs into account.

Figure 1. The cost-effectiveness plane of simulated incremental cost-effectiveness ratios of internet-based cognitive behavioral therapy as an addition to care as usual.

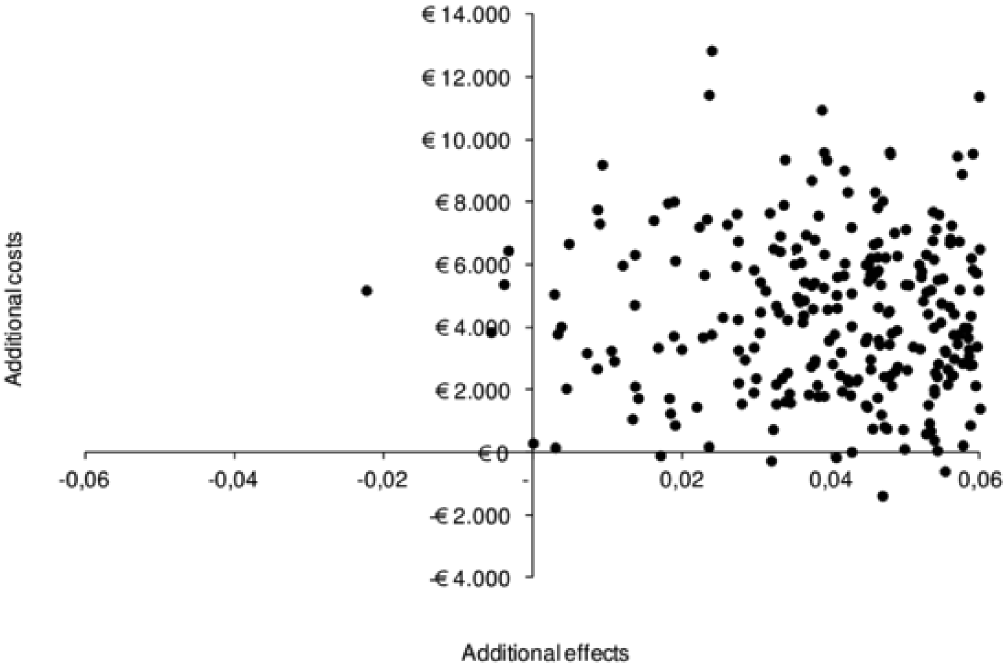


Figure 2. Cost-effectiveness acceptability curve comparing internet-based cognitive behavioral therapy plus care as usual group to care as usual alone group. ICUR: incremental cost-utility ratio; QALY: quality-adjusted life year.

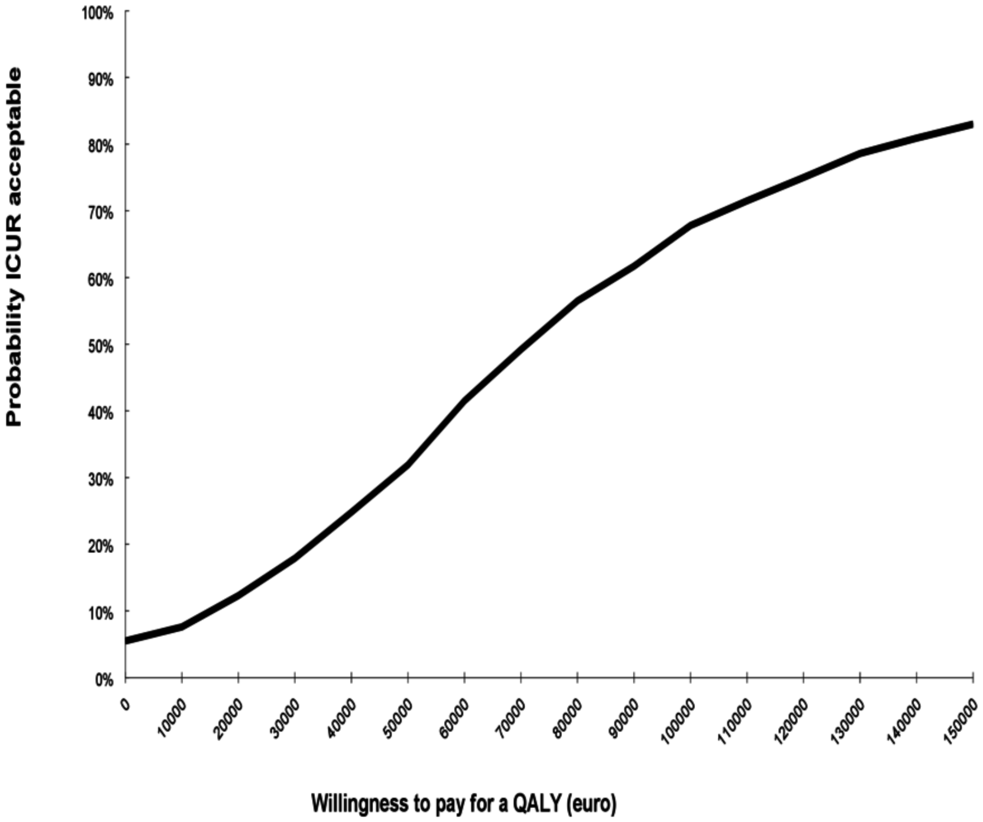


Figure 3. Secondary analysis excluding medication costs: Cost-effectiveness plane of simulated incremental cost-effectiveness ratios of internet-based cognitive behavioral therapy as an addition to care as usual.

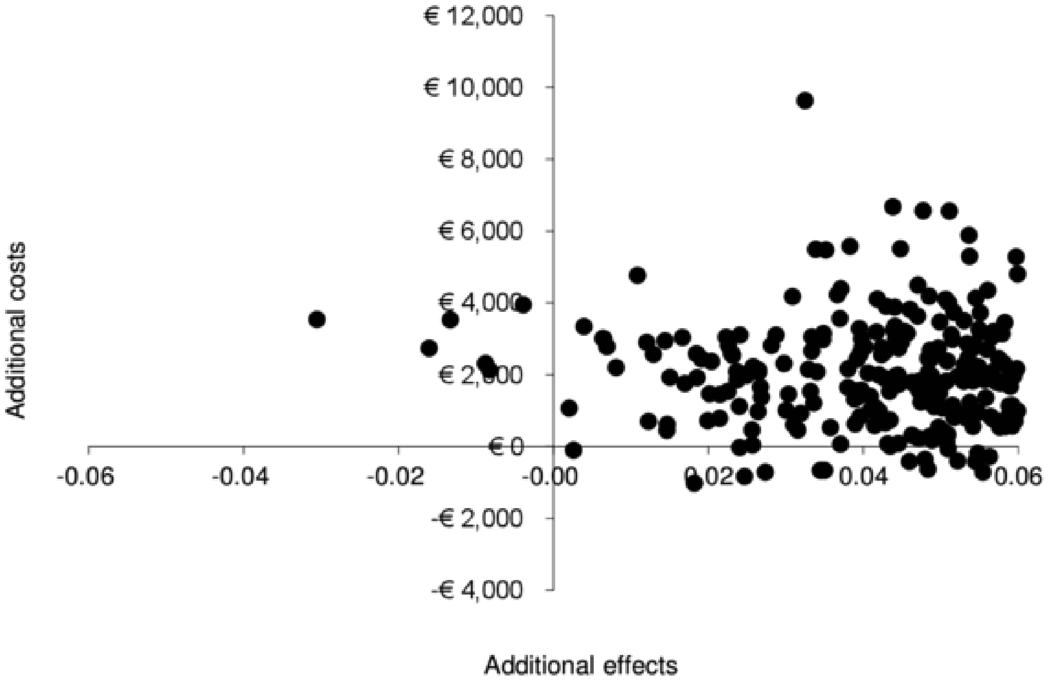
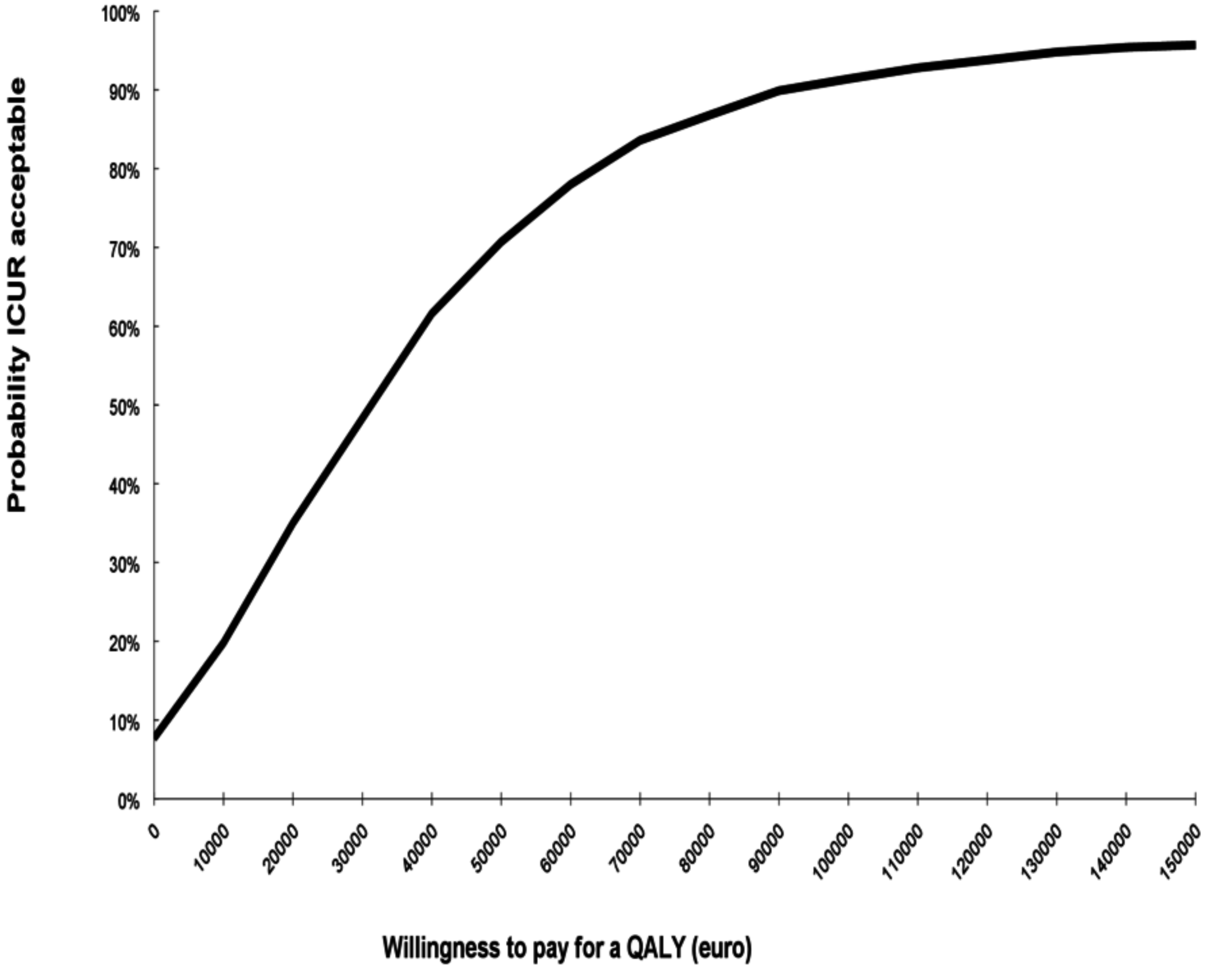


Figure 4. Secondary analysis excluding medication costs: Cost-effectiveness acceptability curve comparing the internet-based cognitive behavioral therapy plus care as usual group to care as usual alone group.



Discussion

This study was conducted with the aim of gaining insight into costs and effects of a tailored, therapist-guided ICBT protocol as an addition to CAU for patients with rheumatoid arthritis and elevated levels of distress. The key findings were (1) a positive effect on QALYs was observed in the intervention group compared with that in the control group; (2) cost ratios showed that this effect came at a greater cost to society; and (3) substantial costs in this population were generated by medications, and there were no group differences with regard to this. The cost-benefit ratio improved when costs of the medications were excluded. Based on effects of the intervention on improving the quality of life, implementation of the intervention is recommended; however, with respect to its effect on the costs, further study is warranted.

Results of this study are in agreement with those of previous studies that reported promising results with regard to the cost-effectiveness of therapist-guided psychological interventions [22,23]. Guidance by a therapist comes at a cost, which accumulates with the duration of treatment. Reducing therapist time by, for example, using more automated or prewritten responses could be beneficial in terms of costs, but comes with the possible risk of losing tailored aspects of the intervention. Future research should compare the cost-effectiveness of a guided, partially-guided, or nonguided intervention to identify the optimal amount of guidance necessary for obtaining cost-effective results. Total costs of the intervention per patient were very low in comparison to the other costs accounted for, which makes the intervention a relatively cheap addition to standard care for patients who might benefit from this intervention in terms of their HRQoL.

Medication costs within the field of rheumatoid arthritis have received considerable attention as biologics have a relatively high cost, and a similar finding was observed in this study. Although in the past there have been some indications that improved psychological functioning increases medication adherence and lowers medication use in the long term [12], no group differences in medication costs were observed in this study. Exclusion of medication costs from the analysis showed a more beneficial cost-effectiveness ratio of the ICBT intervention. Although adherence to medication was not an explicit goal of this intervention, it would be worthwhile to examine the ability of internet-based interventions in changing medication adherence patterns and medication use in patients with rheumatoid arthritis [34]. These interventions could include, for example, motivational interviews aimed at adherence [35]. Additional societal gains could also be attained by finding ways to enable patients in actively participating in the workforce [36,37].

The results of this study should be interpreted with caution because of the presence of missing data, because not all patients filled out all required measurements for the economic evaluation. Although the last observation carried forward method was applied for missing values, missing values can potentially lead to biases in results.

In conclusion, the tailored, therapist-guided ICBT intervention in patients with rheumatoid arthritis with elevated levels of distress rendered higher effects on HRQoL with higher costs. However, society may be willing to pay for the intervention as these costs remain within the threshold generally stated for health care interventions in the Netherlands. The findings of this study were in support of implementation of the intervention as a potential addition to CAU for patients with rheumatoid arthritis with elevated levels of distress, although future studies are necessary for optimizing the cost-benefit ratio.

Acknowledgments

We would like to thank Nina Koch and Milou Looijmans for collection and registration of data; Saskia Spillekom-van Koulil, Tamara Bremer, Els Garritsen, Irene Vermeulen, Lieke Wirken, and Alicia Wijnakker for their work as Web-based therapists; our patient research partners Henk van Duijn, Mariëtte Tomas, Hen Ros, and Ilse van Ee, who contributed to the development of the intervention website and research; and IPPZ for collaboration in designing the Web-based program and for ongoing technical support.

This research was funded by grants from ZonMw (The Netherlands Organisation for Health Research and Development, 80-82310-98-09060) and Pfizer (WS682746). ZonMw and Pfizer were not involved in the study design, data collection, data analysis, manuscript preparation, or publication decisions.

Conflicts of Interest

None declared.

References

1. Evers AWM, Zautra A, Thieme K. Stress and resilience in rheumatic diseases: a review and glimpse into the future. Nat Rev Rheumatol 2011 Jun 21;7(7):409-415 [FREE Full text] [doi: 10.1038/nrrheum.2011.80] [Medline: 21691322]
2. Englbrecht M, Kruckow M, Araujo E, Rech J, Schett G. The interaction of physical function and emotional well-being in rheumatoid arthritis--what is the impact on disease activity and coping? Semin Arthritis Rheum 2013 Apr;42(5):482-491 [FREE Full text] [doi: 10.1016/j.semarthrit.2012.09.003] [Medline: 23369659]
3. Matcham F, Scott IC, Rayner L, Hotopf M, Kingsley GH, Norton S, et al. The impact of rheumatoid arthritis on quality-of-life assessed using the SF-36: a systematic review and meta-analysis. Semin Arthritis Rheum 2014 Oct;44(2):123-130 [FREE Full text] [doi: 10.1016/j.semarthrit.2014.05.001] [Medline: 24973898]
4. Nas K, Sarac AJ, Gur A, Cevik R, Altay Z, Erdal A, et al. Psychological status is associated with health related quality of life in patients with rheumatoid arthritis. J Back Musculoskelet Rehabil 2011;24(2):95-100. [doi: 10.3233/BMR-2011-0281] [Medline: 21558614]
5. Singh JA, Nelson DB, Fink HA, Nichol KL. Health-related quality of life predicts future health care utilization and mortality in veterans with self-reported physician-diagnosed arthritis: the veterans arthritis quality of life study. Semin Arthritis Rheum 2005 Apr;34(5):755-765 [FREE Full text] [doi: 10.1016/j.semarthrit.2004.08.001] [Medline: 15846592]
6. Martikainen JA, Kautiainen H, Rantalaiho V, Puolakka KT. Longterm Work Productivity Costs Due to Absenteeism and Permanent Work Disability in Patients with Early Rheumatoid Arthritis: A Nationwide Register Study of 7831 Patients. J Rheumatol 2016 Dec;43(12):2101-2105 [FREE Full text] [doi: 10.3899/jrheum.160103] [Medline: 27909140]
7. Verstappen SMM. Rheumatoid arthritis and work: The impact of rheumatoid arthritis on absenteeism and presenteeism. Best Pract Res Clin Rheumatol 2015 Jun;29(3):495-511 [FREE Full text] [doi: 10.1016/j.berh.2015.06.001] [Medline: 26612244]
8. Bojke L, Spackman E, Hinde S, Helliwell P. Capturing all of the costs in NICE appraisals: the impact of inflammatory rheumatic diseases on productivity. Rheumatology (Oxford) 2012 Feb;51(2):210-215 [FREE Full text] [doi: 10.1093/rheumatology/ker348] [Medline: 22267153]
9. Husberg M, Davidson T, Hallert E. Non-medical costs during the first year after diagnosis in two cohorts of patients with early rheumatoid arthritis, enrolled 10 years apart. Clin Rheumatol 2017 Mar;36(3):499-506 [FREE Full text] [doi: 10.1007/s10067-016-3470-z] [Medline: 27832385]
10. Hallert E, Husberg M, Kalkan A, Bernfort L. Rheumatoid arthritis is still expensive in the new decade: a comparison between two early RA cohorts, diagnosed 1996-98 and 2006-09. Scand J Rheumatol 2016 Oct;45(5):371-378 [FREE Full text] [doi: 10.3109/03009742.2015.1126344] [Medline: 26854719]

11. Wagner CJ, Metzger FG, Sievers C, Marschall U, L'hoest H, Stollenwerk B, et al. Depression-related treatment and costs in Germany: Do they change with comorbidity? A claims data analysis. J Affect Disord 2016 Mar 15;193:257-266. [doi: 10.1016/j.jad.2015.12.068] [Medline: 26773923]
12. Sharpe L, Allard S, Sensky T. Five-year followup of a cognitive-behavioral intervention for patients with recently-diagnosed rheumatoid arthritis: effects on health care utilization. Arthritis Rheum 2008 Mar 15;59(3):311-316 [FREE Full text] [doi: 10.1002/art.23312] [Medline: 18311768]
13. Astin JA, Beckner W, Soeken K, Hochberg MC, Berman B. Psychological interventions for rheumatoid arthritis: a meta-analysis of randomized controlled trials. Arthritis Rheum 2002 Jun 15;47(3):291-302 [FREE Full text] [doi: 10.1002/art.10416] [Medline: 12115160]
14. Sharpe L. Psychosocial management of chronic pain in patients with rheumatoid arthritis: challenges and solutions. J Pain Res 2016;9:137-146 [FREE Full text] [doi: 10.2147/JPR.S83653] [Medline: 27042139]
15. Englbrecht M, Gossec L, DeLongis A, Scholte-Voshaar M, Sokka T, Kvien TK, et al. The impact of coping strategies on mental and physical well-being in patients with rheumatoid arthritis. Semin Arthritis Rheum 2012 Feb;41(4):545-555 [FREE Full text] [doi: 10.1016/j.semarthrit.2011.07.009] [Medline: 22340997]
16. Ferwerda M, van Beugen S, van Middendorp H, Spillekom-van Koulil S, Donders ART, Visser H, et al. A tailored-guided internet-based cognitive-behavioral intervention for patients with rheumatoid arthritis as an adjunct to standard rheumatological care: results of a randomized controlled trial. Pain 2017 May;158(5):868-878 [FREE Full text] [doi: 10.1097/j.pain.0000000000000845] [Medline: 28106666]
17. Evers AWM, Kraaimaat FW, van Riel PLCM, de Jong AJL. Tailored cognitive-behavioral therapy in early rheumatoid arthritis for patients at risk: a randomized controlled trial. Pain 2002 Nov;100(1-2):141-153. [Medline: 12435467]
18. 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]
19. van Beugen S, Ferwerda M, Hoeve D, Rovers MM, Spillekom-van Koulil S, van Middendorp M, 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]
20. Buhrman M, Gordh T, Andersson G. Internet interventions for chronic pain including headache: A systematic review. Internet Interventions 2016 May;4:17-34. [doi: 10.1016/j.invent.2015.12.001]
21. Tate DF, Finkelstein EA, Khavjou O, Gustafson A. Cost effectiveness of internet interventions: review and recommendations. Ann Behav Med 2009 Aug;38(1):40-45 [FREE Full text] [doi: 10.1007/s12160-009-9131-6] [Medline: 19834778]
22. Hedman E, Ljótsson B, Lindefors N. Cognitive behavior therapy via the Internet: a systematic review of applications, clinical efficacy and cost-effectiveness. Expert Rev Pharmacoecon Outcomes Res 2012 Dec;12(6):745-764. [doi: 10.1586/erp.12.67] [Medline: 23252357]
23. Donker T, Blankers M, Hedman E, Ljótsson B, Petrie K, Christensen H. Economic evaluations of Internet interventions for mental health: a systematic review. Psychol Med 2015 Dec;45(16):3357-3376. [doi: 10.1017/S0033291715001427] [Medline: 26235445]
24. Lorig KR, Ritter PL, Laurent DD, Plant K. The internet-based arthritis self-management program: a one-year randomized trial for patients with arthritis or fibromyalgia. Arthritis Rheum 2008 Jul 15;59(7):1009-1017 [FREE Full text] [doi: 10.1002/art.23817] [Medline: 18576310]
25. Hifinger M, Hiligsmann M, Ramiro S, Watson V, Severens JL, Fautrel B, et al. Economic considerations and patients' preferences affect treatment selection for patients with rheumatoid arthritis: a discrete choice experiment among European rheumatologists. Ann Rheum Dis 2017 Jan;76(1):126-132 [FREE Full text] [doi: 10.1136/annrheumdis-2016-209202] [Medline: 27190098]
26. Felson DT, Anderson JJ, Boers M, Bombardier C, Furst D, Goldsmith C, et al. American College of Rheumatology. Preliminary definition of improvement in rheumatoid arthritis. Arthritis Rheum 1995 Jun;38(6):727-735. [Medline: 7779114]
27. Evers AW, Taal E, Kraaimaat FW, Jacobs JW, Abdel-Nasser A, Rasker JJ, et al. A comparison of two recently developed health status instruments for patients with arthritis: Dutch-AIMS2 and IRGL. Arthritis Impact Measurement Scales. Impact of Rheumatic diseases on General health and Lifestyle. Br J Rheumatol 1998 Feb;37(2):157-164. [Medline: 9569070]
28. Lamers LM, Stalmeier PF, McDonnell J, Krabbe PF, van Busschbach JJ. [Measuring the quality of life in economic evaluations: the Dutch EQ-5D tariff]. Ned Tijdschr Geneeskd 2005 Jul 09;149(28):1574-1578. [Medline: 16038162]
29. Bouwmans C, De Jong K, Timman R, Zijlstra-Vlasveld M, Van der Feltz-Cornelis C, Tan Swan S, et al. Feasibility, reliability and validity of a questionnaire on healthcare consumption and productivity loss in patients with a psychiatric disorder (TiC-P). BMC Health Serv Res 2013;13:217 [FREE Full text] [doi: 10.1186/1472-6963-13-217] [Medline: 23768141]
30. Roijen L, Linden N, Bouwmans C, Kanters T, Tan S. Kostenhandleiding: Methodologie van kostenonderzoek en referentieprijzen voor economische evaluaties in de gezondheidszorg. In: Zorginstituut Nederland. The Netherlands: Zorginstituut Nederland; 2015.
31. Nivel. URL: https://www.nivel.nl/sites/default/files/bestanden/Tabellenboek_Reumatische_aandoeningen_Nederland.pdf [WebCite Cache ID 70fReh9VJ]

32. Koopmanschap MA. PRODISQ: a modular questionnaire on productivity and disease for economic evaluation studies. Expert Rev Pharmacoecon Outcomes Res 2005 Feb;5(1):23-28 [FREE Full text] [doi: 10.1586/14737167.5.1.23] [Medline: 19807557]
33. Raad voor de Volksgezondheid en Zorg. 2006. Zinnige en duurzame zorg (Fair and sustainable care) URL: https://www.raadrvs.nl/uploads/docs/Achtergrondstudie_-_Zicht_op_zinnige_en_duurzame_zorg.pdf [accessed 2018-01-30] [WebCite Cache ID 6wrufj4zp]
34. Morgan C, McBeth J, Cordingley L, Watson K, Hyrich KL, Symmons DPM, et al. The influence of behavioural and psychological factors on medication adherence over time in rheumatoid arthritis patients: a study in the biologics era. Rheumatology (Oxford) 2015 Oct;54(10):1780-1791 [FREE Full text] [doi: 10.1093/rheumatology/kev105] [Medline: 25972390]
35. Georgopoulou S, Prothero L, Lempp H, Galloway J, Sturt J. Motivational interviewing: relevance in the treatment of rheumatoid arthritis? Rheumatology (Oxford) 2016 Dec;55(8):1348-1356 [FREE Full text] [doi: 10.1093/rheumatology/kev379] [Medline: 26515960]
36. Mahalik J, Shigaki CL, Baldwin D, Johnstone B. A review of employability and worksite interventions for persons with rheumatoid arthritis and osteoarthritis. Work 2006;26(3):303-311. [Medline: 16720971]
37. Hoving JL, van Zwieten MC, van der Meer M, Sluiter JK, Frings-Dresen MH. Work participation and arthritis: a systematic overview of challenges, adaptations and opportunities for interventions. Rheumatology (Oxford) 2013 Jul;52(7):1254-1264 [FREE Full text] [doi: 10.1093/rheumatology/ket111] [Medline: 23472042]

Abbreviations

CAU: care as usual
EQ-5D-3L: EuroQol-5dimensions-3levels
HRQoL: health-related quality of life
ICBT: internet-based cognitive behavioral therapy
ICUR: incremental cost-utility ratio
QALY: quality-adjusted life year
RCT: randomized controlled trial
TiC-P: Costs associated with Psychiatric Illness

Edited by G Eysenbach; submitted 01.02.18; peer-reviewed by J Sturt, S Matsuda; comments to author 01.03.18; revised version received 16.04.18; accepted 08.05.18; published 11.10.18.

Please cite as:
Ferwerda M, van Beugen S, van Middendorp H, Visser H, Vonkeman H, Creemers M, van Riel P, Kievit W, Evers A
Tailored, Therapist-Guided Internet-Based Cognitive Behavioral Therapy Compared to Care as Usual for Patients With Rheumatoid Arthritis: Economic Evaluation of a Randomized Controlled Trial
J Med Internet Res 2018;20(10):e260
URL: https://www.jmir.org/2018/10/e260/
doi:10.2196/jmir.9997
PMID:30309835

Original Paper

Eliciting the Impact of Digital Consulting for Young People Living With Long-Term Conditions (LYNC Study): Cognitive Interviews to Assess the Face and Content Validity of Two Patient-Reported Outcome Measures

Jackie Sturt[1], BA (Hons), RN, PhD; Thandiwe Rebecca Dliwayo[2], BSc (Hons), MSc; Vera Forjaz[1], MSc; Kathryn Hamilton[3], BSc (Hons), MSc, PhD; Carol Bryce[4], MA, PhD; Joseph Fraser[5], BA (Hons), MA; Frances Griffiths[4,6], MBBS, MRCGP, PhD

[1]Florence Nightingale Faculty of Nursing, Midwifery and Palliative Care, King's College London, London, United Kingdom
[2]Division of Psychiatry, University College London, London, United Kingdom
[3]Research Department of Clinical, Educational and Health Psychology, University College London, London, United Kingdom
[4]Warwick Medical School, University of Warwick, Coventry, United Kingdom
[5]Joe's Diabetes Ltd, London, United Kingdom
[6]Centre for Health Policy, University of the Witwatersrand, Johannesburg, South Africa

Corresponding Author:
Jackie Sturt, BA (Hons), RN, PhD
Florence Nightingale Faculty of Nursing, Midwifery and Palliative Care
King's College London
James Clerk Maxwell Building
57 Waterloo Road
London, SE18WA
United Kingdom
Phone: 44 07743190301
Email: jackie.sturt@kcl.ac.uk

Abstract

Background: Digital consulting, using email, text, and Skype, is increasingly offered to young people accessing specialist care for long-term conditions. No patient-reported outcome measures (PROMs) have been evaluated for assessing outcomes of digital consulting. Systematic and scoping reviews, alongside patient involvement, revealed 2 candidate PROMs for this purpose: the patient activation measure (PAM) and the physician's humanistic behaviors questionnaire (PHBQ). PAM measures knowledge, beliefs, and skills that enable people to manage their long-term conditions. PHBQ assesses the presence of behaviors that are important to patients in their physician-patient interactions.

Objective: This study aimed to assess the face and content validity of PAM and PHBQ to explore whether they elicit important outcomes of digital consulting and whether the PROMs can isolate the digital consultation component of care.

Methods: Participants were drawn from 5 clinics providing specialist National Health Service care to 16- to 24-year-olds with long-term health conditions participating in the wider LYNC (Long-Term Conditions, Young People, Networked Communications) study. Overall, 14 people undertook a cognitive interview in this substudy. Of these, 7 participants were young people with either inflammatory bowel disease, cystic fibrosis, or cancer. The remaining 7 participants were clinicians who were convenience sampled. These included a clinical psychologist, 2 nurses, 3 consultant physicians, and a community youth worker practicing in cancer, diabetes, cystic fibrosis, and liver disease. Cognitive interviews were transcribed and analyzed, and a spreadsheet recorded the participants' PROM item appraisals. Illustrative quotes were extracted verbatim from the interviews for all participants.

Results: Young people found 11 of the PAM 13 items and 7 of the additional 8 PAM 22 items to be relevant to digital consulting. They were only able to provide spontaneous examples of digital consulting for 50% (11/22) of the items. Of the 7 clinicians, 4 appraised all PAM 13 items and 20 of the PAM 22 items to be relevant to evaluating digital consulting and articulated operationalization of the items with reference to their own digital consulting practice with greater ease than the young people. Appraising the PHBQ, in 14 of the 25 items, two-thirds of the young people's appraisals offered digital consulting examples with ease, suggesting that young people can detect and discern humanistic clinician behaviors via digital as well as face-to-face communication channels. Moreover, 17 of the 25 items were appraised as relevant by the young people. This finding was mirrored in the clinician appraisals. Both young people and the clinicians found the research task complex. Young participants required considerably more researcher prompting to elicit examples related to digital consulting rather than their face-to-face care.

Conclusions: PAM and PHBQ have satisfactory face and content validity for evaluating digital consulting to warrant proceeding to psychometric evaluation. Completion instructions require revision to differentiate between digital and face-to-face consultations.

(J Med Internet Res 2018;20(10):e268) doi:10.2196/jmir.9786

KEYWORDS

communication; referral and consultation; electronic mail; text messaging; chronic disease; young adult; patient-reported outcome measures

Introduction

Clinical Outcomes in Young Adults

In the United Kingdom, 23% of 11- to 15-year-olds report having a long-term illness or disability and of those reporting a long-term condition, 58% take medication [1]. As these young people approach adulthood and take over greater responsibility for their health from their parents, their health outcomes notably decline. In type 1 diabetes, accessing care declines once they have transitioned from pediatric to adult services [2]. The key clinical outcome of importance to their future health, blood glucose control, deteriorates with 1 in 4 young people having blood glucose levels that will lead to complications of diabetes in adulthood [3,4]. In sickle cell disease, during adolescence, regular attendance at outpatient clinics reduces and poorer adherence to medical regimens, in particular penicillin prophylaxis, is reported [5-7]. This is especially problematic because 25% of young adult deaths due to sickle cell are linked to infection and poor compliance with penicillin prophylaxis [7]. In a birth to 18-year-old cohort with sickle cell, highest mortality occurred in those over 15 years with the transfer from pediatric to adult care settings. This represents a particularly high-risk situation [8]. In asthma, median rates of medication adherence in adolescents have been reported to be 43% [9], and when they present for emergency visits, they are less likely to continue treatment plans compared with younger children [10].

Digital Consulting

In an attempt to improve poor health outcomes, specialist clinical teams caring for young people with long-term conditions have begun to use digital consulting, which commonly consists of email and text messaging [11]. This practice was investigated in the LYNC (Long-Term Conditions, Young People, Networked Communications) study that aimed to understand the role of digital consulting in specialist long-term condition clinics for 16- to 24-year-olds [12]. The LYNC study broadly defined digital consulting as two-way patient-clinician communications for clinical purposes, synchronous or asynchronous, in which either party is, or could be, mobile. The objective of this study was to explore how engagement in self-care is impacted by the use of digital consulting including effects, costs, and necessary safety and ethical safeguards. This was undertaken by interviewing and observing the practice of 173 clinical team members and 165 young people across 20 clinics providing care for young adults with 13 different physical and mental health conditions [11,13]. Young people and their clinicians value digital consulting as providing access to health care in a timely way and at a point when it makes a difference to how the young people manage their health [11].

Patient-Reported Outcomes

Although these digital methods are increasingly being used by health professionals and are supported by health-provider organizations [14-16], their impact on patient-reported health outcomes has not been explored. In planning and undertaking health services improvement or experimental research, patient-reported outcomes (PROs) are of documented importance, and where you have a PRO, you need a patient-reported outcome measure (PROM) [17]. Examples of PROs are treatment satisfaction, depressive symptoms, quality of life, and pain.

Systematic reviews of digital consulting relevant to young people with long-term conditions identify gaps in our understanding of relevant PROs [12]. Digital consulting PROs need to demonstrate some amenability to change because of care being provided in both a face-to-face and digital contexts. For example, would the PRO be influenced if the young person had routine hospital appointments supplemented by email contact with one of their clinical teams between routine appointments? We might not expect depressive symptoms to change, but we might anticipate that treatment satisfaction could perhaps be influenced. Systematic reviews have also recommended undertaking research across several clinical populations using digital consulting [12]. To achieve this, a PRO that can span different long-term conditions and elicit generic outcomes related to the potential benefits of digital consulting is required.

Identifying Candidate Patient-Reported Outcome Measures for Digital Consulting

As part of the LYNC study, a series of systematic and scoping reviews were undertaken alongside patient involvement to identify generic PROs and PROMs that assessed the impact of types of patient-clinician interactions (not specifically digital interactions) on patient engagement with their health as an outcome. This work is published elsewhere [13], but in summary, an initial systematic review to identify existing candidate generic PROMs to be used for evaluating digital consulting was undertaken. The review generated 28 generic PROMs from which we expected to identify an existing PROM that would be fit for purpose. To make the public and patient engagement task of appraising these PROMs manageable, they were clustered in PRO themes. The main PRO clusters were

(1) quality of life (x 8 PROMs), (2) psychological status (x 5 PROMs), and (3) patient satisfaction with information technology (x 4 PROMs). Results of a Web-based public engagement survey (n=57) and discussion with the LYNC study public engagement team indicated that the PRO clusters were outdated and not fit for the purpose of evaluating the impact of digital consulting for young adults. Service user engagement in determining important outcomes of care is essential, and these discussions within the LYNC public engagement team were informed by the early LYNC fieldwork. This resulted in a decision to undertake a scoping review to identify interventions (and their PROMs) that aimed to improve face-to-face hospital appointments primarily by increasing patient participation in them. Two reviews were identified [18,19], and with reference to the emerging LYNC fieldwork data, the Harrington et al [18] review was particularly useful in informing a second Web-based public engagement survey to gain perspectives on the most important attributes of a successful face-to-face medical appointment. This survey (n=143) resulted in the emergence of 4 priorities. Using these as our guide, we returned to the Harrington et al review [18] to identify the PROMs that had been used to assess the impacts of interventions used in routine appointments to promote the following: (1) question asking, (2) engagement between the clinician and patient, (3) recall following an appointment, and (4) having a sense of disease control. None of the studies in the Harrington et al review [18] had utilized PROMs to assess these attributes specifically, but the scoping review, the public engagement process, and early LYNC fieldwork informed us about what we needed to look for in the PROM literature. A final scoping review identified 5 candidate PROMs that the LYNC public engagement team narrowed down to 3, which appeared potentially fit for the purpose of eliciting patient outcomes targeted by digital consulting and suitable for use across multiple long-term conditions. These PROMs were (1) the patient activation measure (PAM 22) [20], (2) the patient activation measure short form (PAM 13) [21], and (3) the physician's humanistic behaviors questionnaire (PHBQ) [22]. The 13 items of the short form PAM 13 are contained within the 22 items of the PAM 22. Therefore, items from just 2 candidate PROMs, the PAM 22 and the PHBQ, were the focus of our interest.

The Patient Activation Measure 22 and the Physician's Humanistic Behaviors Questionnaire

The aim of PAM is to understand the knowledge, beliefs, and skills required by people to enable them to manage their long-term conditions. PAM 22 demonstrates construct and criterion validity [20]. PAM 13, a short-form version of PAM 22, has been more extensively examined for use in additional populations and languages [21]. It offers strong psychometric evidence in people living with long-term conditions [23-34]. PAM, specifically PAM-13, possesses a large body of evidence that supports its strong psychometric properties across diverse cultures and patient groups. PAM 13 and 22 are proprietorial PROMs that can limit access. The aim of PHBQ instrument development was to understand what physician humanistic behaviors, performed (or not) in the physician-patient interaction, were important to patients and to develop a measure to assess these behaviors in different health care contexts [22]. No further evaluations of the PHBQ full scale were identified, although a number of investigators have used items from PHBQ in new outcome measures or used PHBQ as a benchmark for convergent validity when developing new measures [35-37]. In summary, PHBQ evidences satisfactory face, content, and convergent validity but reports no reliability evidence. Research endeavor is required to address its psychometric properties relating to reliability and validity in other clinical contexts. A full assessment of the psychometric validation evidence for both PROMs is provided in Multimedia Appendix 1.

Study Objectives

Neither PAM nor PHBQ have been used to evaluate the impact of digital consulting, and their face and content validity for this purpose have not been established. Our research objective was to assess the face and content validity of these measures for assessing the patient-clinician relationship impacts of digital consulting. In addition, where PROM items had face validity with young people and their clinicians, we also aimed to access whether they were thinking about the digital consultation component of their care when reflecting on their activation and their patient-clinician relationship.

Methods

Research Design

We undertook a qualitative exploration of PHBQ and PAM using cognitive interviewing, an established method for the development and revision of PROMs [38-42]. Criteria for PROM development and revision are that the items have clarity, relevance, and are unambiguous for their purpose [38]. Cognitive interviews use a concurrent interviewing approach where participants are asked to *think aloud* as they consider the measure or a debriefing approach where participants are asked to consider their experience of answering the questions and completing the measure immediately after its completion [38,39]. We used both concurrent and debriefing approaches in all interviews to understand the cognitions of the participants as they were considering individual items and their overall experience of each PROM and its fitness for purpose in the use of digital consulting. We acknowledged that there may be organizational barriers for clinicians in delivering care aligned with the individual PROM items in a digital care context. Consequently, it was important to expose each PROM to clinicians in addition to young people. This would identify whether there are differences in perspectives between people receiving digitally enhanced care and those offering it.

Setting

A total of 5 clinics from the wider LYNC study participated in this substudy. LYNC study clinics spanned England and Wales and provided specialist National Health Service (NHS) care to 16- to 24-year-olds with long-term mental or physical health conditions [11,13]. The majority of clinicians and participants in the main LYNC study clinics were engaging in digital consultations in addition to routine clinic appointments.

Participants

Interviewees were a subset of LYNC study participants with experience of digital consulting [12]. A total of 7 young people aged 16 to 24 years with a diagnosis of inflammatory bowel disease (4), cystic fibrosis (1), or cancer (2) consented to have a cognitive interview from a total sample of 165 young people recruited to the LYNC study. A total of 7 clinicians from 5 NHS trusts, also from the LYNC study pool of 173 clinical team members, participated. Of these, 1 was a clinical psychologist, 2 were nurses, 3 were consultant physicians, and 1 was a community youth worker. They worked in the fields of cancer (2), diabetes (2), cystic fibrosis (1), and liver disease (2). Convenience sampling was used in 5 of the recruited LYNC clinical sites. The complexity of the cognitive interviewing task meant that we needed young adult participants to agree to participate in a face-to-face interview. Everyone who was approached agreed to participate in the cognitive interview. We obtained ethical approval (14/WM/0066) from the National Research Ethics Service Committee West Midlands—The Black Country.

Procedure

Three experienced researchers completed the interviews. Clinician participants were briefed on the purpose of the interview and the rationale for each candidate PROM. They were then asked to consider each item and articulate what they were considering as they read it. As each PROM was aimed at patient completion and not clinician completion, the clinicians were asked to consider the extent to which they themselves performed or supported the behaviors and attitudes in the items when thinking about digital consulting. They were asked to consider their own clinical care using the response options included in each measure (eg, strongly agree to strongly disagree). In doing so, they would be prompted to consider the item's relevance in a digital consulting context. A total of 5 clinicians reviewed both PROMs and 2 reviewed 1 PROM each. Young people's PROM cognitive interviews took place at the end of a shortened regular LYNC interview [11]. The young participants were briefed on the purpose of the interview and of the measure. To reduce participant burden, each young person was asked to consider, in rotation, 1 in 3 of the items in both measures. The sequential rotation of shortened questionnaires was documented to ensure item coverage during interviews. Young people were primed when considering the measures to think about their current access to, and communication with, their health care team, including digital consultations, and how completely they felt the questions addressed the sorts of things that are important to them in terms of engaging in digital consultations. We expected that young people would find it difficult to separate out their face-to-face interactions from their digitally enhanced care. Consequently, young people were asked to *think aloud* as they considered the items in the measure in relation to their own care and articulate a personal response to each item, and where necessary, they were prompted to clarify whether they were appraising in terms of their experience of digital consultations. Both clinicians and patients were also prompted to reflect on the measures in their entirety after having reviewed the individual items in terms of the overall relevance of the items for assessing the outcomes of digital consultations and whether there were any important aspects of this assessment that were currently omitted. All interviews were audio recorded, transcribed verbatim, and anonymized.

Analysis

For both young person and clinician interviews, a spreadsheet was developed to indicate, by item, whether the participant could appraise the item as related to digital consulting with ease, operationalized by offering examples as they *thought aloud*, and whether the item was considered relevant. Illustrative quotes were extracted verbatim from the transcribed interviews for all participants. Where 2 or more young people and 4 or more clinicians appraised the item as relevant, we determined the item to be relevant.

Results

Face and Content Validity Assessments of the Patient Activation Measure 13 and 22 Items

PAM 22 was reviewed in entirety by 6 of the 7 clinicians, and selected items were reviewed in rotation by 7 young people. Each individual item of PAM 22 was reviewed a minimum of 8 times overall. Multimedia Appendix 2 presents the degree to which young peoples' and clinicians' responses indicated that they agreed or strongly agreed with the relevance of each item as an outcome of digital consulting. It also presents the degree to which digitally enhanced care was referred to spontaneously with ease or when prompted while considering the relevance of the item operationalized as examples being given in the interviews. Illustrative quotes from the cognitive interviews show where digital consulting was shown to be underpinning the construct of patient activation.

Young People

Multimedia Appendix 2 illustrates that in comparison with clinicians, the young people were less able to appraise the items as related to digital consulting with ease. In only 50% (11/22) of the items was more than 1 young person, out of the 2 or 3 reviewing the items, able to give an illustrative *think aloud* example. This did not differ according to whether they were appraising PAM 13 or PAM 22 items. Young people had less difficulty with appraising item relevance for digital consulting. On PAM 13, 11 of the items were assessed as being relevant to young peoples' experience of managing their health condition with digital access to clinicians. Of the additional 8 PAM 22 items, 7 were found to be relevant. In articulating digital examples, the young people provided a similar proportion of these across all of the PAM 13 and PAM 22 items identified to be relevant. We had anticipated that young people would struggle to critique the measure items and instead use the item as a mechanism for appraising their abilities or their care. In appraising both the PAM 13 and 22 items, we found that young people did default to this position while undertaking this complex task. For example, in the last quote in Multimedia Appendix 2, the young person offers evidence of how they demonstrate that they can handle symptoms at home rather than on whether being able to handle their symptoms at home may be facilitated by the use of digital consulting. One young person reported that some of the questions were similar to others.

Another suggested that PAM 22 could be used to assess and compare whether young people preferred digital versus face-to-face communication. The complex nature of the task limited the extent to which the young people were able to critique the PAM scale overall in depth and comment on item duplication or item omissions in important areas related to digital consulting.

Clinicians

Clinician participants had an easier grasp of the task than the young people and were able to appraise the items with relative ease. At least 4 of the 6 clinicians were able to address the items as related to digital consulting with ease on all PAM 13 items and 20 of the PAM 22 items. The majority of clinicians agreed that the measure was suitable for appraising digital consulting. For example, on PAM 13, if all 6 clinicians had agreed with the relevance of all 13 items, a 100% agreement rate would have been established by a total agreement score of 78. The agreement score was 63, which indicates high levels of agreement across the items and the clinicians on the item relevance for evaluating digital consulting. Most were able to provide examples of the relevance of the items to digitally enhanced care in support of their assessment. The clinicians were more able to critique the PAM items beyond the direct task and raised a number of issues. In relation to some items on the PAM 22 scale, with item similarities observed in the following paired questions—item 1 and 2; 13 and 16; and 18 and 20—made either item redundant. A diabetes consultant commented that they struggled to establish the relevance of item 3 and 6 in relation to evaluating outcomes for digital consulting and suggested that further clarification was required. The same clinician also remarked about the significant overlap between item 10, 12, and other items in the instrument. A liver clinician also found items 15 and 20 to be ambiguous and therefore could not provide a response.

Two clinicians suggested that there were items that may be more or less applicable for particular conditions. For example, a clinician suggested that item 18 on PAM 22 was condition-specific and would yield different responses in various health conditions. The young people's community worker in cancer services disagreed with the idea in item 1 inferring that the patient should take sole responsibility for their health condition and reported that it took a joint effort to maintain good health outcomes for each patient. The diabetes consultant reported that they were yet to encounter a patient who had successfully made changes in adopting a healthy lifestyle when considering item 16 on the PAM 22 scale. A young people's community worker in cancer services commented that the measure appeared sufficiently comprehensive by covering a broad context of how technology might contribute to supporting self-management and found items 4 and 8 to be particularly relevant. However, this health professional added that not all items of PAM 22 would be relevant to all patient groups. This idea was supported by a cystic fibrosis consultant who reported that although they strongly agreed with the first 6 items on the scale, some items relating to taking sole action of one's health condition would be less important for a cystic fibrosis patient group than for other health conditions. The clinician reported that due to the nature of the health condition, young people living with cystic fibrosis were usually knowledgeable about managing their health condition at home and infrequently visited their general physicians in the community.

A lead cancer nurse summarized that PAM 22 adequately probed the right issues in terms of assessing the outcomes of digital consulting. This idea was also supported by a diabetes consultant, but he/she concluded that although the instrument was exhaustive, it was quite repetitive. Some concepts were missing for some clinicians in confirming the PAM 13 or 22 as a comprehensive measure of the effects of digital consulting. The liver clinician stated that it was paramount for any assessment to include items that would gauge whether there were notable changes in young people's confidence attributed to digital consulting. Overall, PAM was appraised as having a satisfactory level of face and content validity notwithstanding the issues raised regarding some of the items.

Face and Content Validity Assessments of the Physician Humanistic Behaviors Questionnaire

Assessments were made by 6 of the 7 clinicians and 7 young people, and each individual item was assessed a minimum of 8 times. Multimedia Appendix 3 presents data of item relevance and illustrative examples where digital consulting was identified to be underpinning the constructs and/or outcomes of physician humanistic behaviors.

Young People

Compared with the clinicians, the 7 young people appraising these items found it very difficult to engage in a lengthy critique. Young people had a tendency to answer the question rather than articulate its relevance to digital consulting. The young people required prompting to clarify whether their responses were underpinned by digital consulting, and on some occasions, they could not elaborate any further beyond one-word responses when prompted about this. Nonetheless, Multimedia Appendix 3 illustrates that for 56% (13/25) of the items, at least 2 of the 3 young people reviewing each item were able to offer digital examples with ease. This suggests that young people can detect and discern humanistic clinician behaviors via digital as well as face-to-face communication channels. In terms of item relevance, 68% (17/25) of the items were apprised as relevant by 2 or more young people. As with the PAM appraisal, the complexity of the task limited a wider critique on item repetition or omission by the young people.

Clinicians

If all 6 clinician appraisals had agreed with the relevance of all 25 items, a 100% agreement rate would have been established, resulting in a total agreement score of 150. The agreement score was 108, which indicates reasonable levels of agreement across the items and the clinicians on the item relevance for evaluating digital consulting. Clinicians were able to provide examples of how the items elicited aspects of their performance relating to digital consulting. There were queries raised with regard to the ambiguity and applicability of some items on PHBQ. They were concerned that the idea of making promises to patients in item 7 was uncomfortable and suggested that the item needed readjustment. A diabetes clinician reported that item 3, *is in a hurry*, required a context to be given to make the question clear for the patients to respond. Moreover, 3 clinicians had queries regarding the relevance of item 21. The professionals stated that they were already familiar with their patients and would know how they preferred to be addressed. One clinician agreed with the relevance of item 1, which probed clinicians' ability to follow through on patients' problems and provided examples of how this aspect of performance related to digital consulting. Item 11 was thought to be irrelevant with regard to evaluating digital consulting by the same health professional, and this idea was also supported by a cystic fibrosis consultant who expressed a similar opinion with regard to item 16. These 2 professionals considered item 17 to be inappropriate for a digital consulting scenario, although 1 clinician appreciated that abruptness could be communicated in an email.

A small number of recommendations were made to either add or alter existing items on the scale. Two clinicians suggested that the addition of items that probed whether clinicians respond to queries in a timely manner would be valuable. An individual expressed concern that PHBQ appeared to scrutinize clinicians' behaviors without much consideration of young people's conduct toward health professionals. They suggested that the instrument needed to account for the prior agreement established between health professionals and young people with respect to methods of communication and essentially probe those dynamics. They concluded that PHBQ was useful in asking the right questions to obtain young people's opinions about the outcome of digital consulting in terms of the clinician's behavior.

Discussion

Principal Findings

Cognitive interviews with 7 clinicians and 7 young people found that PAM and PHBQ contain items that were satisfactorily appraised by clinicians and patients as having reasonable degrees of face and content validity. In addition, both scales have demonstrated themselves to be credible clinical evaluation measures for assessing the effects of digital consulting. The young people found it difficult to separate their digital experiences from their face-to-face consultation experiences. The clinicians had less difficulty with this task and context. Each young person's experience is limited by the n-of-1 nature of their clinical experiences, whereas the clinicians have multiple young patients upon whom they were drawing to produce illustrative examples with ease.

Strengths and Limitations of the Research Methods

The cognitive interviews provided in-depth analyses of 2 PROMs with both clinicians and patients. We recruited a broad range of health professional disciplines to participate, and they were experienced providers of digitally enhanced care. Consequently, they were evaluating the measures' real clinical relevance for use in a service evaluation or audit context. In relation to the young people, they were all receivers of digitally enhanced care, and all patients evaluated some aspects of both measures. These participants were conveniently sampled, and their representativeness of the whole LYNC study population cannot be assured. Despite 4 of the main LYNC case study sites being mental health services, none of the cognitive interviews were undertaken with clinicians or young people from a mental health service, and this is a limitation of the evidence presented.

The most apparent methodological weakness is that most young people and some clinicians did not understand the exercise of having a cognitive interview and the additional challenge of distinguishing their digital and face-to-face consulting experiences. The interviewers had to work hard with the time and level of understanding that each participant presented. In this respect, it could be argued that some interviews were not sufficiently accessing young people's, autonomous cognitions on the topic. Often, considerable prompting from the researcher was required to facilitate the young person's understanding of the interview purpose and elicit the required data. The PHBQ scale was developed for the purpose of appraising physician behaviors, and we applied it to a broader range of health professionals. Despite this original focus, no participant referred to any limitation of this broader application.

Strengths and Limitations of the Patient-Reported Outcome Measures

Previous research has demonstrated that psychometrically, PAM 13 has stronger and more detailed all-round validity in demonstrating clear face, content, construct, and criterion validity and psychometric reliability across populations [21,23-34]. A large body of evidence, therefore, demonstrates that PAM-13 is a valid, reliable, and clinically useful measure of patient activation that can be used across diverse cultures and patient groups. In light of the evidence presented above, the strong psychometric properties of PAM 13, in particular, support its use in adults with long-term conditions. Its validity in a digital context has not previously been established in the literature. PAM 22 has been psychometrically appraised fewer times [20]. The PAM 22 and 13 outcomes in young adult populations have been reported [43-45], although Disabato et al [45] questioned the appropriateness of PAM 13 in this population. Its psychometric evaluation in a young adult German population with a mean age of 17.5 years has however since been satisfactorily established [46]. The PHBQ has less documented detail about the face and content validity for face-to-face consultations in a single study appraising its psychometric properties [22]. A revised version of PHBQ was used in a study of 1400 service men and women in which 41% of the study population were aged 18 to 24 years [47]. The authors reported no methodological limitation in the use of the measure with their participants.

During interviews, clinician participants were not specifically directed to appraise both PROMs as research tools and in their principal roles as clinicians; they may have adopted a view of them as clinical tools. This may have contributed to some negative appraisals regarding item or topic repetition within the scale. Asking the same question in different ways is often a strength in a research tool to enable the research team to tease out response congruity. In a clinical setting, where time for both clinicians and patients is at a premium, repetition may appear wasteful and irritating. Our LYNC findings point to the role of digital consulting as having particular strength when there is a pre-existing relationship between the patient and the clinician [11]. This aligns well with the PHBQ scale, which is assessing physician humanistic behaviors, where 2 previous face-to-face consultations are required before scale completion.

Comparisons With Prior Work

PAM is one of the few PROMs that serves twin purposes of performing well as both a clinical and research tool. Armstrong et al (2017) evaluated the use of PAM as a clinical tool in patient care and found it to prompt a range of both positive and negative responses by both patients and clinicians and concluded that further evaluative work in its clinical application is warranted [48]. PROMs are increasingly important across the health care sector. The US Food and Drug Administration (FDA) [49] has produced guidance on their use by the pharmaceutical industry in recognition that the evolution of science and health care products and practices should not come at any price. There is concern that drugs can have detrimental effects on the constructs measured by PROMs, and it is important to identify this early in the evaluation of the drug so that it can be optimized to reduce patient treatment burden, for example, on quality of life. Although this advice is specifically intended for the pharmaceutical industry, Speight and Barendse [17] remind us that the principles underpinning the FDA's guidance apply across other contexts. Health care is awash with technological solutions, and in the area of digital consulting, NHSmail2, a secure email service for sharing patient identifiable and sensitive information, is currently undergoing a rollout across the NHS with the specific objective of promoting digital communication between many users, including patients. With such fundamental shifts in the way that health care will be delivered and received on the horizon, it is imperative that we ensure we have clinically and scientifically fit-for-purpose PROMs available that have credibility with patients, clinicians, researchers, and technological innovators.

Recommendations for Research

The strength of PAM 13 evidence indicates that it represents the best-evidenced PROM for evaluating the impact of digital consulting on patient activation behaviors. Our research indicates that revision of the completion instructions would be necessary to enable participants or patients to differentiate between digital and face-to-face consultation in their responses. Subsequent research studies need to expose PAM 13, 22, and PHBQ to clinical evaluations of digitally enhanced services to examine how they perform psychometrically in a larger population. Face and content validity of these PROMs need to be explored in mental health settings delivering digitally enhanced care. Although we acknowledge that these PROMs may not be perfect and may benefit from item addition, revision, and removal, they are presented as the best PROMs available at the current time to evaluate the impact of digital consulting for young adults with long-term conditions. The development of a new PROM for this purpose may be appropriate in the future.

Acknowledgments

The authors would like to thank the LYNC project research administrators, Morna Woods and Rebecca Rogers; the LYNC project management group; and in particular, the following patient and public involvement members: Sandra Atkinson, Clare Martin, Nidah Chaudhry, Joseph Eaton, Bijal Trivedi, and Della Maria Heck for supporting the patient-reported outcomes prioritization activity that preceded the empirical work presented. This project was funded by the National Institute of Health Research, Health Services and Delivery Research Programme (project number 12/209/51).

Authors' Contributions

JS and FG conceived and led the study. VF, KH, and CB devised the interview schedules and collected the data. TRD and JS analyzed the data and drafted the manuscript. All authors contributed to the manuscript.

Conflicts of Interest

FG received grants from the UK National Institute of Health Research during the conduct of the study. JF received personal fees from NHS England, outside the submitted work.

Multimedia Appendix 1

Strength of validity and reliability evidence for the candidate patient-reported outcome measures.

[PDF File (Adobe PDF File), 56KB - jmir_v20i10e268_app1.pdf]

Multimedia Appendix 2

Clinician and young people's face and content validity assessments of the Patient Activation Measure 13 and 22 items.

[PDF File (Adobe PDF File), 68KB - jmir_v20i10e268_app2.pdf]

Multimedia Appendix 3

Clinicians and young people's face and content validity assessments of the Physician Humanistic Behavior Questionnaire.

[PDF File (Adobe PDF File), 58KB - jmir_v20i10e268_app3.pdf]

References

1. Brooks F, Magnusson J, Klemera E, Chester K, Spencer N, Smeeton N. HBSC England National Report: Health Behaviour in School-aged Children HBSC: World Health Organization Collaborative Cross National Study. Hatfield, Hertfordshire: University of Hertfordshire; 2015.
2. Royal College of Paediatric and Child Health. London; 2016. National Paediatric Diabetes Audit 2014-2015 URL: https://www.rcpch.ac.uk/sites/default/files/NPDA_Annual_Report_for_2015-16_Care_Processes_and_Outcomes.pdf [WebCite Cache ID 72P26cTp8]
3. Royal College of Paediatrics and Child Health. London; 2012. National Paediatric Diabetes Audit Report 2010-11 URL: https://www.rcpch.ac.uk/sites/default/files/2018-03/npda_national_report_2010-11.pdf [WebCite Cache ID 72Ofj5nZd]
4. Digibete. 2016. National Diabetes Transition Audit, 2003-14 England and Wales URL: https://www.digibete.org/wp-content/uploads/2017/11/National_Diabetes_Transition_Audit_Report_2003_2014.pdf [accessed 2018-09-13] [WebCite Cache ID 72P5v83C0]
5. Witherspoon D, Drotar D. Correlates of adherence to prophylactic penicillin therapy in children with sickle cell disease. Child Health Care 2010 Jun 07;35(4):281-296. [doi: 10.1207/s15326888chc3504_1]
6. Howard J, Woodhead T, Musumadi L, Martell A, Inusa BP. Moving young people with sickle cell disease from paediatric to adult services. Br J Hosp Med (Lond) 2010 Jun;71(6):310-314. [doi: 10.12968/hmed.2010.71.6.48435] [Medline: 20551868]
7. Musumadi L, Westerdale N, Appleby H. An overview of the effects of sickle cell disease in adolescents. Nurs Stand 2012;26(26):35-40. [doi: 10.7748/ns2012.02.26.26.35.c8973] [Medline: 22482187]
8. Quinn CT, Rogers ZR, McCavit TL, Buchanan GR. Improved survival of children and adolescents with sickle cell disease. Blood 2010 Apr 29;115(17):3447-3452 [FREE Full text] [doi: 10.1182/blood-2009-07-233700] [Medline: 20194891]
9. Naimi DR, Freedman TG, Ginsburg KR, Bogen D, Rand CS, Apter AJ. Adolescents and asthma: why bother with our meds? J Allergy Clin Immunol 2009 Jun;123(6):1335-1341 [FREE Full text] [doi: 10.1016/j.jaci.2009.02.022] [Medline: 19395075]
10. Cooper WO, Hickson GB. Corticosteroid prescription filling for children covered by Medicaid following an emergency department visit or a hospitalization for asthma. Arch Pediatr Adolesc Med 2001 Oct;155(10):1111-1115. [Medline: 11576005]
11. Griffiths F, Bryce C, Cave J, Dritsaki M, Fraser J, Hamilton K, et al. Timely digital patient-clinician communication in specialist clinical services for young people: a mixed-methods study (The LYNC study). J Med Internet Res 2017 Apr 10;19(4):e102 [FREE Full text] [doi: 10.2196/jmir.7154] [Medline: 28396301]
12. Griffiths FE, Atherton H, Barker JR, Cave JA, Dennick K, Dowdall P, et al. Improving health outcomes for young people with long term conditions: the role of digital communication in current and future patient-clinician communication for NHS providers of specialist clinical services for young people - LYNC study protocol. Digit Health 2015;1:2055207615593698 [FREE Full text] [doi: 10.1177/2055207615593698] [Medline: 29942543]
13. Griffiths FE, Armoiry X, Atherton H, Bryce C, Buckle A, Cave JA, et al. The role of digital communication in patient-clinician communication for NHS providers of specialist clinical services for young people The LYNC study: a mixed methods study. NIHR Journals Library 2018;6(9) [FREE Full text] [doi: 10.3310/hsdr06090] [Medline: 29485818]
14. Department of Health. 2012. The power of information: putting all of us in control of the health and care information we need URL: https://data.gov.uk/sites/default/files/DH%20Open%20Data%20Strategy_10.pdf [WebCite Cache ID 72Oj78lAI]
15. Wachter R. The Wachter Review of Health IT: Update and Insights. In: The King's Fund. 2016 Presented at: Digital Health and Care Congress; July 5-6, 2016; The Kings' Fund, London.
16. Wachter RM. Making IT work: harnessing the power of health information technology to improve care in England. London, England: England NAGoHITi; 2016 Sep 07. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/550866/Wachter_Review_Accessible.pdf [accessed 2018-09-13] [WebCite Cache ID 72Ojo8YAU]
17. Speight J, Barendse SM. FDA guidance on patient reported outcomes. BMJ 2010 Jun 21;340:c2921. [doi: 10.1136/bmj.c2921]
18. Harrington J, Noble LM, Newman SP. Improving patients' communication with doctors: a systematic review of intervention studies. Patient Educ Couns 2004;52(1):7-16. [doi: 10.1016/S0738-3991(03)00017-X] [Medline: 14729285]
19. Pendleton D, Hasler J. Doctor-Patient Communication. Cambridge, MA: Academic Press Inc; 1983.
20. Hibbard JH, Stockard J, Mahoney ER, Tusler M. Development of the Patient Activation Measure (PAM): conceptualizing and measuring activation in patients and consumers. Health Serv Res 2004 Aug;39(4 Pt 1):1005-1026 [FREE Full text] [doi: 10.1111/j.1475-6773.2004.00269.x] [Medline: 15230939]
21. Hibbard JH, Mahoney ER, Stockard J, Tusler M. Development and testing of a short form of the patient activation measure. Health Serv Res 2005 Dec;40(6 Pt 1):1918-1930 [FREE Full text] [doi: 10.1111/j.1475-6773.2005.00438.x] [Medline: 16336556]
22. Weaver MJ, Ow CL, Walker DJ, Degenhardt EF. A questionnaire for patients' evaluations of their physicians' humanistic behaviors. J Gen Intern Med 1993 Mar;8(3):135-139. [doi: 10.1007/BF02599758] [Medline: 8455109]

23. Rademakers J, Nijman J, van der Hoek L, Heijmans M, Rijken M. Measuring patient activation in The Netherlands: translation and validation of the American short form Patient Activation Measure (PAM13). BMC Public Health 2012 Jul 31;12:577 [FREE Full text] [doi: 10.1186/1471-2458-12-577] [Medline: 22849664]
24. Skolasky RL, Green AF, Scharfstein D, Boult C, Reider L, Wegener ST. Psychometric properties of the patient activation measure among multimorbid older adults. Health Serv Res 2011 Apr;46(2):457-478 [FREE Full text] [doi: 10.1111/j.1475-6773.2010.01210.x] [Medline: 21091470]
25. Zill JM, Dwinger S, Kriston L, Rohenkohl A, Härter M, Dirmaier J. Psychometric evaluation of the German version of the Patient Activation Measure (PAM13). BMC Public Health 2013 Oct 30;13:1027 [FREE Full text] [doi: 10.1186/1471-2458-13-1027] [Medline: 24172020]
26. Schmaderer M, Pozehl B, Hertzog M, Zimmerman L. Psychometric properties of the Patient Activation Measure in multimorbid hospitalized patients. J Nurs Meas 2015;23(3):128-141. [doi: 10.1891/1061-3749.23.3.E128] [Medline: 26673761]
27. Green CA, Perrin NA, Polen MR, Leo MC, Hibbard JH, Tusler M. Development of the Patient Activation Measure for Mental Health. Adm Policy Ment Health 2010 Jul;37(4):327-333 [FREE Full text] [doi: 10.1007/s10488-009-0239-6] [Medline: 19728074]
28. Moljord IE, Lara-Cabrera ML, Perestelo-Pérez L, Rivero-Santana A, Eriksen L, Linaker OM. Psychometric properties of the Patient Activation Measure-13 among out-patients waiting for mental health treatment: A validation study in Norway. Patient Educ Couns 2015 Nov;98(11):1410-1417 [FREE Full text] [doi: 10.1016/j.pec.2015.06.009] [Medline: 26146239]
29. Maindal HT, Sokolowski I, Vedsted P. Translation, adaptation and validation of the American short form Patient Activation Measure (PAM13) in a Danish version. BMC Public Health 2009 Jun 29;9:209 [FREE Full text] [doi: 10.1186/1471-2458-9-209] [Medline: 19563630]
30. Ahn YH, Yi CH, Ham OK, Kim BJ. Psychometric properties of the Korean version of the Patient Activation Measure 13 (PAM13-K) in patients with osteoarthritis. Eval Health Prof 2015 Jun;38(2):255-264. [doi: 10.1177/0163278714540915] [Medline: 24986844]
31. Stepleman L, Rutter MC, Hibbard J, Johns L, Wright D, Hughes M. Validation of the patient activation measure in a multiple sclerosis clinic sample and implications for care. Disabil Rehabil 2010;32(19):1558-1567. [doi: 10.3109/09638280903567885] [Medline: 20590506]
32. Packer TL, Kephart G, Ghahari S, Audulv Å, Versnel J, Warner G. The Patient Activation Measure: a validation study in a neurological population. Qual Life Res 2015 Jul;24(7):1587-1596. [doi: 10.1007/s11136-014-0908-0] [Medline: 25557496]
33. Hung M, Carter M, Hayden C, Dzierzon R, Morales J, Snow L, et al. Psychometric assessment of the patient activation measure short form (PAM-13) in rural settings. Qual Life Res 2013 Apr;22(3):521-529. [doi: 10.1007/s11136-012-0168-9] [Medline: 22466721]
34. Brenk-Franz K, Hibbard J, Herrmann W, Freund T, Szecsenyi J, Djalali S, et al. Validation of the German version of the patient activation measure 13 (PAM13-D) in an international multicentre study of primary care patients. PLoS One 2013;8(9):e74786 [FREE Full text] [doi: 10.1371/journal.pone.0074786] [Medline: 24098669]
35. Glaser KM, Markham FW, Adler HM, McManus PR, Hojat M. Relationships between scores on the Jefferson Scale of physician empathy, patient perceptions of physician empathy, and humanistic approaches to patient care: a validity study. Med Sci Monit 2007 Jul;13(7):CR291-CR294. [Medline: 17599021]
36. Dine CJ, Ruffolo S, Lapin J, Shea JA, Kogan JR. Feasibility and validation of real-time patient evaluations of internal medicine interns' communication and professionalism skills. J Grad Med Educ 2014 Mar;6(1):71-77 [FREE Full text] [doi: 10.4300/JGME-D-13-00173.1] [Medline: 24701314]
37. She M, Li Z, Rau PL. Physician Communication Behaviors that Predict Patient Trust in Outpatient Departments. : Springer; 2015 Presented at: International Conference on Cross-Cultural Design; 2015; Crete, Greece.
38. Ahmed N, Bestall JC, Payne SA, Noble B, Ahmedzai SH. The use of cognitive interviewing methodology in the design and testing of a screening tool for supportive and palliative care needs. Support Care Cancer 2009 Jun;17(6):665-673. [doi: 10.1007/s00520-008-0521-2] [Medline: 18982363]
39. Drennan J. Cognitive interviewing: verbal data in the design and pretesting of questionnaires. J Adv Nurs 2003 Apr;42(1):57-63. [doi: 10.1046/j.1365-2648.2003.02579.x] [Medline: 12641812]
40. Rosal MC, Carbone ET, Goins KV. Use of cognitive interviewing to adapt measurement instruments for low-literate Hispanics. Diabetes Educ 2003;29(6):1006-1017. [doi: 10.1177/014572170302900611] [Medline: 14692374]
41. Knafl K, Deatrick J, Gallo A, Holcombe G, Bakitas M, Dixon J, et al. The analysis and interpretation of cognitive interviews for instrument development. Res Nurs Health 2007 Apr;30(2):224-234. [doi: 10.1002/nur.20195] [Medline: 17380524]
42. Cooke D, O'Hara MC, Beinart N, Heller S, La Marca R, Byrne M, UK NIHR DAFNE Study Group. Linguistic and psychometric validation of the Diabetes-Specific Quality-of-Life Scale in UK English for adults with type 1 diabetes. Diabetes Care 2013 May;36(5):1117-1125 [FREE Full text] [doi: 10.2337/dc12-0076] [Medline: 23250797]
43. Flink M, Lindblad M, Frykholm O, Kneck Å, Nilsen P, Årestedt K, et al. The Supporting Patient Activation in Transition to Home (sPATH) intervention: a study protocol of a randomised controlled trial using motivational interviewing to decrease re-hospitalisation for patients with COPD or heart failure. BMJ Open 2017 Jul 10;7(7):e014178 [FREE Full text] [doi: 10.1136/bmjopen-2016-014178] [Medline: 28698319]

44. Huang JS, Terrones L, Tompane T, Dillon L, Pian M, Gottschalk M, et al. Preparing adolescents with chronic disease for transition to adult care: a technology program. Pediatrics 2014 Jun;133(6):e1639-e1646 [FREE Full text] [doi: 10.1542/peds.2013-2830] [Medline: 24843066]
45. Disabato JA, Cook PF, Hutton L, Dinkel T, Levisohn PM. Transition from pediatric to adult specialty care for adolescents and young adults with refractory epilepsy: a quality improvement approach. J Pediatr Nurs 2015;30(5):e37-e45. [doi: 10.1016/j.pedn.2015.06.014] [Medline: 26260696]
46. Bomba F, Markwart H, Mühlan H, Menrath I, Ernst G, Thyen U, et al. Adaptation and validation of the German Patient Activation Measure for adolescents with chronic conditions in transitional care: PAM 13 for Adolescents. Res Nurs Health 2018 Dec;41(1):78-87. [doi: 10.1002/nur.21831] [Medline: 29266283]
47. Douglas SR, Vides de Andrade AR, Boyd S, Leslie M, Webb L, Davis L, et al. Communication training improves patient-centered provider behavior and screening for soldiers' mental health concerns. Patient Educ Couns 2016 Dec;99(7):1203-1212. [doi: 10.1016/j.pec.2016.01.018] [Medline: 26884315]
48. Armstrong N, Tarrant C, Martin G, Manktelow B, Brewster L, Chew S. Independent evaluation of the feasibility of using the Patient Activation Measure in the NHS in England-Final report. Leicester, UK: University of Leicester and The Health Foundation; 2017 Apr. URL: https://lra.le.ac.uk/bitstream/2381/40449/2/PAM%20learning%20set_final%20evaluation%20report_final.pdf [WebCite Cache ID 72OrZBdXV]
49. Patrick DL, Burke LB, Powers JH, Scott JA, Rock EP, Dawisha S, et al. Patient-reported outcomes to support medical product labeling claims: FDA perspective. Value Health 2007;10 Suppl 2:S125-S137 [FREE Full text] [doi: 10.1111/j.1524-4733.2007.00275.x] [Medline: 17995471]

Abbreviations

FDA: Food and Drug Administration
NHS: National Health Service
PAM: patient activation measure
PHBQ: physician’s humanistic behaviors questionnaire
PRO: patient-reported outcome
PROM: patient-reported outcome measure

Edited by G Eysenbach; submitted 02.02.18; peer-reviewed by D Cooke, J Pecina; comments to author 22.02.18; revised version received 30.05.18; accepted 18.06.18; published 11.10.18.

Please cite as:
Sturt J, Dliwayo TR, Forjaz V, Hamilton K, Bryce C, Fraser J, Griffiths F
Eliciting the Impact of Digital Consulting for Young People Living With Long-Term Conditions (LYNC Study): Cognitive Interviews to Assess the Face and Content Validity of Two Patient-Reported Outcome Measures
J Med Internet Res 2018;20(10):e268
URL: https://www.jmir.org/2018/10/e268/
doi:10.2196/jmir.9786
PMID:30309832

Original Paper

Responding to Unsolicited Medical Requests from Health Care Professionals on Pharmaceutical Industry-Owned Social Media Sites: Three Pilot Studies

Andrea M TenBarge[1*], PharmD; Jennifer L Riggins[1*], PharmD, FDIA

Medical Digital Strategy and Capabilities, Eli Lilly and Company, Indianapolis, IN, United States

[*]all authors contributed equally

Corresponding Author:
Andrea M TenBarge, PharmD
Medical Digital Strategy and Capabilities
Eli Lilly and Company
Lilly Corporate Center
Indianapolis, IN,
United States
Phone: 1 317 276 4320
Email: tenbarge_andrea_marie@lilly.com

Abstract

Background: The use of social media has risen tremendously over the past decade with usage rates spanning from 5% American adults in 2005 up to 69% in 2016. A 2011 survey of 4033 clinicians found that 65% physicians use social media for professional purposes. To meet the changing needs and preferences of their customers, medical information departments within the pharmaceutical industry must continue to assess new digital channels such as social media and evolve their medical information services.

Objective: The objective of the study was to pilot the use of social media as an additional channel to respond to unsolicited medical requests from health care professionals (HCP RUR) directed toward the pharmaceutical industry.

Methods: From November 2016 to June 2017, 3 pilots were conducted during 3 professional congresses: the 2016 American College of Rheumatology Annual Meeting, the 2017 American Society of Clinical Oncology Annual Meeting, and the 2017 American Headache Society Annual Scientific Meeting. For each social media account, an identified community manager monitored the incoming account feed for proper triaging of posts. When an unsolicited medical request appeared, the community manager routed the question to the Tier One medical information contact center agents to respond. The following metrics were collected: total number of unsolicited requests directed to medical information contact center agents, total number of unsolicited requests that required escalation to Tier Two medical information associates, total number of unsolicited requests that were confirmed US HCPs, total number of unsolicited requests received after hours, and total number of unsolicited requests that were redirected to a different channel.

Results: During the 3 pilots, 9 unsolicited medical requests were received with request numbers ranging from 2 to 4 requests per pilot. Of these, 1 was from a confirmed US HCP that required escalation to the Tier Two medical information associates. A majority of requests (7 out of 9) came in after the scheduled monitoring hours. There were 4 requests redirected to the medical information contact center phone number. The marketing accounts received more unsolicited medical requests than the corporate accounts (7 vs 2, respectively), and the 3 Twitter accounts saw more overall engagement (ie, medical requests and other general engagement) than the LinkedIn account.

Conclusions: A limited number of medical questions were asked by confirmed HCPs using social media during the 3 pilots. New innovative medical information contact center channels often take time to build awareness. Continued channel awareness is needed to fully understand the channel's desired use. Because consumers currently make up a majority of social media engagement, companies should look into creating a combined consumer and HCP RUR strategy to provide a better experience for all customers.

(J Med Internet Res 2018;20(10):e285) doi:10.2196/jmir.9643

KEYWORDS

social media; health care professionals; medical information

Introduction

The Evolution of Medical Information

Medical affairs organizations within the pharmaceutical industry act as the connecting link between research & development and commercialization teams. A core responsibility of Medical Affairs is to provide up-to-date, accurate, balanced, and nonpromotional responses to unsolicited medical requests regarding the company's products. The action of responding to unsolicited requests (RUR) is performed by the Medical Affairs function, commonly referred to as medical information [1]. Over the past few decades, medical information departments have needed to evolve continually to meet the ever-changing digital landscape. Response channels used for RUR have ranged from basic channels such as the phone, fax, and hard copy in the 1980s and 1990s to the evolution of email responses and self-service websites in the 2000s [2]. Today, the world of digital channels has grown exponentially, leading to a focus on omnichannel response fulfillment with growing possibilities such as Web-based chat, texting, videoconferencing, podcasts, mobile apps, artificial intelligence, and social media. To meet the changing needs and preferences of their customers, medical information departments must continue to assess new digital channels for RUR, such as social media, and evolve their medical information services [2-3].

Social Media Use and Health Care

The use of social media has risen tremendously over the past decade with usage rates spanning from only 5% of American adults in 2005 up to 69% in 2016. The most prevalent users continue to be the younger population with 86% of 18-29-year-olds using social media, followed by 80% of 30-49-year-olds, 64% of 50-64-year-olds, and 34% of those aged 65+ years [4]. A 2012 survey by PricewaterhouseCoopers' Health Research Institute found that 1 in 3 consumers use social media for matters regarding their health with Facebook and YouTube being at the top. A majority of engagement comes from the younger population with 90% of 18-24-year-olds being likely to engage on social media or trust information for health matters compared with 56% of those aged 45-64 years [5].

An analysis of Twitter usage found over 100,000 health care professionals (HCPs) on Twitter in 2014, averaging 295,872 tweets per day. A majority of these HCPs were in the United States (45%), followed by Europe (22%), Near and Far East (17%), South America (13%), and Oceania (2%) [6]. A 2011 survey of 4033 clinicians found that 65% physicians use social media for professional purposes [7]. Furthermore, a 2014 survey of 350 nurse practitioners and physician assistants found that 45% of these HCPs also use social media for business purposes [8]. A search on LinkedIn in November 2017 detected over 400,000 profiles using the term "pharmacist," over 450,000 profiles using the term "physician," and over 1,647,000 profiles using the term "nurse" [9]. Instances of professional use include networking, crowdsourcing, sharing and consuming information, curating information, educating the public, patient engagement and feedback, and discussing health care policy [10-14].

Furthermore, Twitter has increasingly been used during medical conferences to share and discuss information related to the conference [10,13]. Ample literature has been published looking at the role of Twitter at various conferences across specialties [15-22]. A tweet analysis of 13 conferences from 2011-2013 identified 51,159 tweets by 8778 Twitter account holders of which 25% were identified HCPs who composed 19,503 (38%) of the tweets [15].

Social Media and the Pharmaceutical Industry

In recent years, pharmaceutical companies have joined the social media revolution, creating their own corporate or marketing accounts [23-25]. A study of 15 of the top pharmaceutical companies from October 1, 2013 to September 30, 2014 found that 93% of the companies had a company-owned Twitter page, followed by 66% for YouTube, 66% for Facebook, and 60% for LinkedIn. These pages averaged approximately 45,000 subscribers or followers, and a majority of the posts were related to company news (63.4%) and help-seeking electronic direct-to-consumer advertising (40.7%) [23]. In 2014, the Food and Drug Administration (FDA) put out 3 draft guidance documents for industry to assist companies in developing their strategy for participation in social media [26-28].

In the 2011 FDA draft guidance on RUR for off-label information, the FDA states that because other responders in public forums often do not have the most up-to-date and robust information that pharmaceutical companies do on their products, it may be in the public's best interest for the company to respond to unsolicited requests regarding their products in these channels [29]. Most pharmaceutical companies, however, have yet to take full advantage of the customer service capabilities that the medical side of the organization can offer on social media [25]. Those that have ventured into this area often do so by redirecting the customer to the company's contact center number when an unsolicited request for information is received [30]. There are 2 companies that have attempted to incorporate medical information services into their social media strategy by partnering with Sermo to allow physicians to submit unsolicited medical requests to the company [31]. Sermo is a global physician-only social media platform that includes over 800,000 physicians from 150 countries around the world [32]. Survey results showed that 73% of the Sermo physicians found direct access to the company's medical expert as valuable or extremely valuable [31].

Social Media Business Case Development

Currently, medical information services are provided through a number of traditional channels such as telephone, email, live chat, and a self-service website. When evaluating the addition of innovative digital channels to the omnichannel nature of medical information services, there is often a lack of information in the primary literature to assist pharmaceutical companies in developing their strategy. Therefore, to fully assess the channel's potential desired use, proof of concept tactics such as market research are frequently utilized before moving forward with a pilot.

From September 2, 2016 to October 3, 2016, a third-party vendor was contracted to initiate, draft, and complete a market research survey analyzing HCP use of social media for medical information. The Web-based survey was completed by 100 HCPs consisting of 50 physicians and 50 allied HCPs. Survey results showed that a majority of physicians (41/50, 82%) and allied HCPs (45/50, 90%) had asked a medical or product-related question on social media before. Furthermore, a majority of physicians (32/50, 64%) and allied HCPs (36/50, 72%) feel that social media is somewhat, very, or extremely valuable for interacting with pharma. Therefore, with the market research providing a positive business case for piloting HCP RUR on social media, the decision was made to move forward to the pilot stage. The objective of the following research is to pilot the use of social media as an additional channel to respond to unsolicited medical requests from HCPs (HCP RUR) directed toward the pharmaceutical industry.

Methods

Pilot Overviews

From November 2016 to June 2017, 3 pilots were conducted covering 3 professional congresses: the 2016 American College of Rheumatology Annual Meeting, the 2017 American Society of Clinical Oncology Annual Meeting, and the 2017 American Headache Society Annual Scientific Meeting (Table 1). Identified social media accounts were limited to those that were company-owned and had HCPs as one of their target audiences. Corporate communications- and therapeutic area-focused accounts that were active during the congresses were prioritized. The social media account owners were required to have a statement on the page that let the audience know how the company would engage on the site. The account owners were also required to have a process in place for handling and reporting adverse events and product complaints, a process for handling misinformation, a process for crisis management, and a triage process in place to ensure requests were directed to the appropriate groups for responding. Engagement tool selection was based on the current engagement tool that the account owner was using. The 2 engagement tools utilized were Spredfast, a social marketing platform (headquartered in Austin, TX), and Social Studio, a product of the Salesforce Marketing Cloud (headquartered in San Francisco, CA).

Process for Responding to Unsolicited Medical Requests from Health Care Professionals

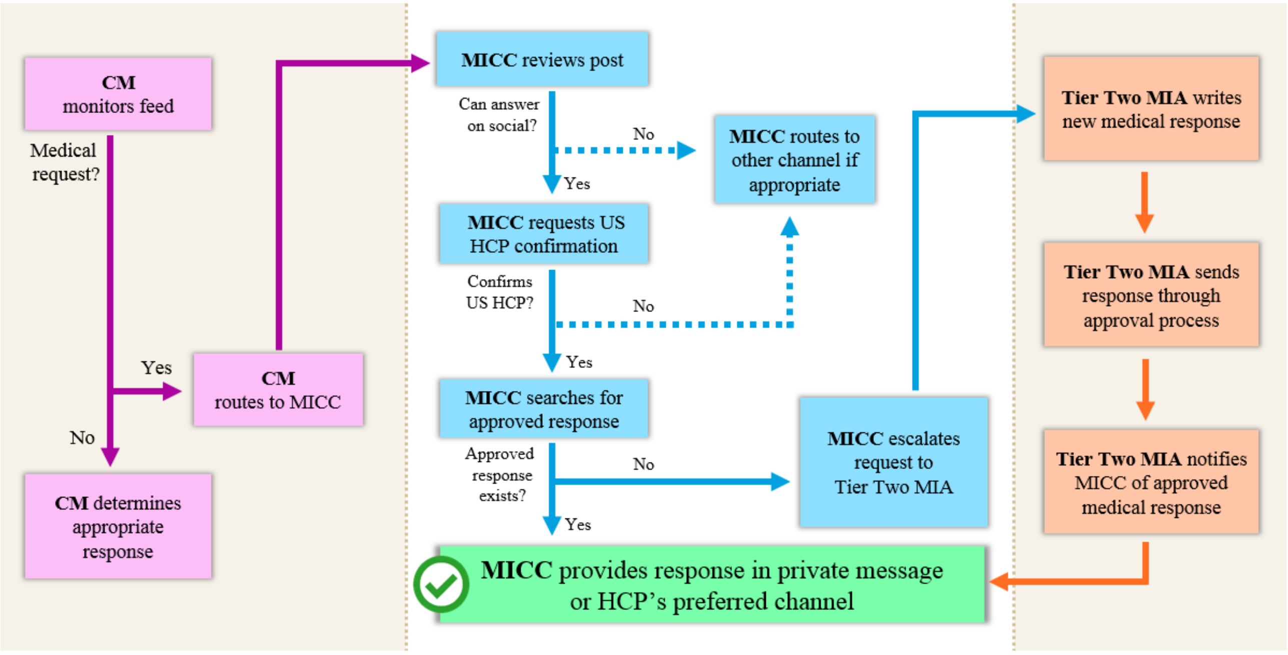
For each social media account, an identified community manager monitored the incoming account feed for proper triaging of posts (Figure 1). When an unsolicited medical request was directed to the company-owned social media account, the community manager would route the question to the Tier One medical information contact center agents to respond. If the question was deemed answerable with a medical response, the medical information contact center agents would direct the requestor to a private message and ask for confirmation that the requestor was a US HCP. If the requestors confirmed they were US HCPs, the medical information contact center agents would then provide the answers within a private message conversation. If the appropriate medical answer was available on the company’s medical information website, medical information contact center agents would provide links to the medical answer within the private message response. This ensured the answers would always remain current and would contain appropriate regulatory information. If an appropriate medical answer was not currently available, the medical information contact center agents would escalate the question to the Tier Two medical information associates for creation and approval of a medical response to answer the question. Due to licensing fees and the limited timeframes of Pilots 1 and 2, a Medical Digital Strategy Consultant acted as a liaison between the medical information contact center agents and the Spredfast engagement tool to ensure proper posting of responses.

In preparation for the pilots, training and simulation sessions were held to teach key participants the process and guardrails for HCP RUR on social media. Subject matter experts were on call to assist with strategy-related questions that occurred during the pilot, including subject matter experts for medical social media operations, contact center operations, social media account operations, and engagement tool operations.

Table 1. Overview of pilots.

Features	Pilot 1	Pilot 2	Pilot 3
Congress	2016 American College of Rheumatology Annual Meeting	2017 American Society of Clinical Oncology Annual Meeting	2017 American Headache Society Annual Scientific Meeting
Pilot dates	November 11-16, 2016	June 2-6, 2017	June 1, 2017-ongoing
Pilot hours	9 am-5 pm EST	9 am-8 pm EST	Ongoing: 9 am-5 pm EST weekdays; Congress: 9 am-8 pm EST
Social accounts	US Rheumatology Twitter handle	Main corporate Twitter handle and US Oncology LinkedIn account	US Migraine Twitter handle
Account owner	Marketing	Corporate	Marketing
Engagement tool	Spredfast	Spredfast	Social Studio

Figure 1. Process for responding to unsolicited medical requests from health care professionals. CM: community manager; MICC: medical information contact center; HCP: health care professional; MIA: medical information associate.



Pilot Metrics

To evaluate the pilots and inform shared learning for future process improvements, the following metrics were identified and collected:

- total number of unsolicited requests directed to medical information contact center agents
- total number of unsolicited requests that required escalation to Tier Two medical information associates
- total number of unsolicited requests that were confirmed US HCPs
- total number of unsolicited requests received after hours
- total number of unsolicited requests that were redirected to a different channel

Results

Pilot 1

During Pilot 1, 4 unsolicited medical requests were received (Table 2 and Figure 2). Of these, 1 was a US HCP asking about registration for a congress event that was to occur the following day. This request required escalation to the Tier Two medical information associate to draft and approve a custom response. Due to the extra effort needed, a response was not able to be provided until shortly before the event occurred. Another was a consumer asking about pipeline information for a specific indication. This requestor was not a US HCP, so an approved pleasantry was used as the response. Even though the question could not be answered, the consumer provided positive feedback, thanking the team for listening. The third question asked for clarification regarding a scientific term in one of the posts. This question was answered correctly by an external Twitter user before the team could respond. The fourth question asked about disease control. However, the requestor never responded to the tweet asking for US HCP confirmation. Of the 4 unsolicited requests, 3 were received after designated monitoring hours.

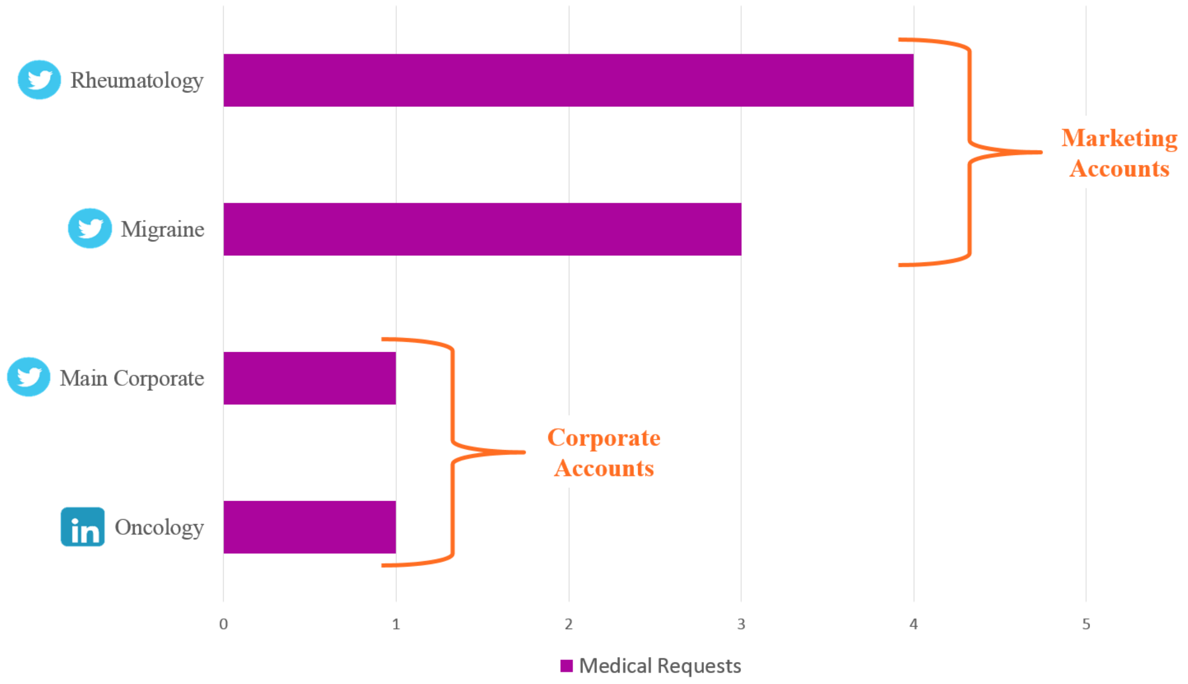
Pilot 2

During Pilot 2, 2 unsolicited medical requests were received (Table 2 and Figure 2). Of these, 1 request was received on the US Oncology LinkedIn account and inquired about drug targeting. Due to the inability to private message between individuals and company pages on LinkedIn, the requestor was directed to the medical information contact center phone number. The second request was received after designated monitoring hours on the company's main corporate Twitter handle. The requestor asked about medication cost and was directed to the medical information contact center phone number owing to the complexity of the response.

Table 2. Medical request metrics by pilot.

Metrics	Pilot 1, n	Pilot 2, n	Pilot 3, n	Total, n
Total medical requests	4	2	3	9
Tier two requests	1	0	0	1
Confirmed US health care professional requests	1	0	0	1
After hours requests	3	1	3	7
Requests redirected to call number	0	2	2	4

Figure 2. Total medical requests by social media account.



Pilot 3

During the first 3 weeks of Pilot 3, 3 unsolicited medical requests were received (Table 2 and Figure 2). Of these, 2 were directed to the medical information contact center phone number (1 asking about nondrug treatment and the other asking about general disease state help). The third question asked about diagnostic imaging. A tweet was sent asking for US HCP confirmation, but the requestor did not reply. All 3 requests came in after designated monitoring hours. No unsolicited medical requests were received during the American Headache Society Annual Scientific Meeting.

Discussion

Pilot Comparison of Request Volume

During the 3 pilots, 9 unsolicited medical requests were received, with medical request numbers ranging from 2 to 4 requests per pilot (Table 2). Pilot 1, during the 2016 American College of Rheumatology Annual Meeting, saw the most activity with 4 total medical requests and 1 Tier Two request. Pilot 2, during the 2017 American Society of Clinical Oncology Annual Meeting, saw the least activity with only 2 total medical requests. All 3 of the requests received during Pilot 3 came in during the first month.

Account Comparison of Request Volume

Of the 4 social media accounts used during the 3 pilots, 2 were owned by Marketing, and 2 were corporate-owned accounts (Figure 2). The marketing accounts received more unsolicited medical requests than the corporate accounts (7 vs 2, respectively). This could be owing to the marketing accounts posting more scientifically detailed material while the corporate accounts posted higher level, corporate-focused information. The 3 Twitter accounts saw more overall engagement (ie, medical requests and other general engagement) than the LinkedIn account, which only received 1 medical request. This may be owing to most of the pilots occurring during a medical congress, where Twitter activity tends to increase with the use of congress-specific hashtags.

Hours of Operation

Normal medical information contact center hours of operation are from 9 am-8 pm EST Monday through Friday. However, the 9 am-5 pm EST monitoring timeframe was chosen for Pilot 1 owing to the scheduled working hours for the community managers. During Pilot 1, 3 of the 4 unsolicited medical requests that were received were after the scheduled monitoring hours of 9 am-5 pm EST. As a result, hours of operation were expanded during Pilots 2 and 3 to cover 9 am-8 pm EST during congress days, when activity was expected to be higher while maintaining the 9 am-5 pm EST working hours during standard business weekdays. Even with the expanded hours, a majority of the requests continued to come in after scheduled monitoring hours during Pilots 2 and 3.

Health Care Professional Versus Consumer Requests

Only 1 of the 9 unsolicited medical requests was from a confirmed US HCP. The other requests were from potential consumers or did not respond to the tweet asking them to confirm they were a US HCP. Although the capability was intentionally set up for US HCPs, much of the engagement on the sites appeared to be from consumers. Having a parallel capability setup for answering medical requests from US consumers in addition to US HCPs would provide a better experience for all customers.

Social Media Question Nuances and Preparation

With the addition of new digital channels comes the potential for new types of questions owing to the nature of the channels. Since HCP use of social media, especially Twitter, tends to increase around medical congresses, congress logistics questions may be more prominent. These questions can be more time-sensitive, as experienced in the pilot, as opposed to questions received in other HCP RUR channels. Having approved responses prepared for standard congress logistics questions would help minimize the response time to the customer while also minimizing the workload of the medical information personnel. Another item to consider when preparing for potential questions is the type of information that is planned to be presented or released during the congress, as this information is more likely to appear in social media discussions.

Awareness of the Capability

When adding new innovative medical information contact center digital channels, awareness is key. Awareness often takes months to years to build up enough responsiveness to see the true desired use of a channel. As was found in the market research proof of concept, a majority of HCPs, who said they have never asked a medical or product-related question on social media, said it was because they simply had not thought to ask this type of question on social media. However, 71% (10/14) of these nonusers said they would consider using social media to ask a medical information question to a pharmaceutical company in the future. Although most retail companies in the consumer market have a large customer service representation on social media, the pharmaceutical industry has been slow to adopt this capability owing to the highly regulated nature of their environment. Many pharmaceutical companies simply refer requestors to their contact center phone number, if they respond at all. Therefore, HCPs may not think their question will be answered if they submit a request to pharma on social media. However, to aid in awareness of the social media pilots, engagement guidelines (including items such as hours of operation) were published on participating social media account pages through a pinned tweet or within the guidelines section of the page. The new social media channel capability was also included within a larger medical-wide awareness campaign, which provided awareness regarding the medical organization and the various channels HCPs may use to ask a medical question.

Limitations

During Pilots 1 and 2, medical information contact center agents did not have direct access to the social media engagement tools owing to the limited timeframe of the pilots. This, in turn, added extra steps to the RUR workflow, which increased the overall response time to the requestor. In addition, pilot metrics included information to aid in pilot strategy and uptake but were limited in helping to gauge the overall response quality. Satisfaction surveys can be critical to help strategy leaders fully understand the quality of the capability. Although the pilots offer great insights into the incorporation of HCP RUR on Twitter, every social media platform tends to have its specific uses and features; for example, LinkedIn does not allow private messaging between company pages and individuals. Additional research needs to be conducted to understand the differences in how a company can respond using various platforms such as Facebook, Sermo, Doximity, etc. Furthermore, these 3 pilots were US-focused. Because regulations, digital channel usage, and digital platforms often vary by country, further research is needed to better understand the nuances in setting up social media RUR capabilities in countries around the world.

Conclusions

The 3 HCP RUR social media pilots produced 9 unsolicited medical requests. While this number is considered low for more traditional channels such as the phone, the new innovative medical information contact center channels often take time to build up an increased level of awareness of the new service offering. Providing channel awareness is critical to fully understand the channel’s true desired use. In addition, because consumers currently appear to make up a majority of social media engagement, companies should look into creating a combined consumer and HCP RUR strategy that can be carried out consistently across sites. In conclusion, the pilots provided pertinent strategy insights such as the initial volume to expect and resources required when developing a social media strategy for HCP RUR.

Conflicts of Interest

Both authors are employees of Eli Lilly and Company.

References

1. Crowley-Nowick P, Smith J. BioProcess International. 2013 Apr 01. The Role of Medical Affairs in Moving from R&D to Commercialization URL: http://www.bioprocessintl.com/analytical/qa-qc/the-role-of-medical-affairs-in-moving-from-randd-to-commercialization-341871/ [accessed 2017-11-21] [WebCite Cache ID 6v8pwFqqb]
2. Fierro L, Savulich D, Fraser C, Gažo A, Wyble C. Historical perspective of technologies used in medical communications, past, present, and future. Drug Inf J 2009 Nov;43(6):705-712. [doi: 10.1177/009286150904300608]
3. Giffin SA, Baumeister P, Bowers D. Review of the evolution of medical information in light of changes in the external landscape. Ther Innov Regul Sci 2014 Nov;48(6):NP6-NP14. [doi: 10.1177/2168479014547935] [Medline: 30227466]
4. Pew Research Center. 2017 Jan 12. Social Media Fact Sheet URL: http://www.pewinternet.org/fact-sheet/social-media/ [accessed 2017-11-21] [WebCite Cache ID 6v8qAeJCy]

5. PwC Health Research Institute. 2012 Apr. Social media likes healthcare: from marketing to social business URL: https://www.pwc.com/us/en/health-industries/health-research-institute/publications/pdf/health-care-social-media-report.pdf [accessed 2017-11-21] [WebCite Cache ID 6v8qLGUwZ]
6. Lester J. Creation. 2014 Sep. 100,000 healthcare professionals on Twitter-study wins DIA poster competition URL: http://creation.co/knowledge/healthcare-professionals-on-twitter-study-wins-competition/ [accessed 2017-11-21] [WebCite Cache ID 6v8qOqJZK]
7. Modahl M, Tompsett L, Moorhead T. QuantiaMD. 2011 Sep. Doctors, patients & social media URL: http://www.quantiamd.com/q-qcp/social_media.pdf [accessed 2017-11-21] [WebCite Cache ID 6v98WuOrP]
8. Blazek N. Clinical Advisor. 2014. 2014 Social Media Survey URL: http://www.clinicaladvisor.com/2014-social-media-survey/slideshow/2223/ [accessed 2017-11-21] [WebCite Cache ID 6v99SYoZJ]
9. LinkedIn. 2017 Nov. Profile Search URL: https://www.linkedin.com/ [accessed 2017-11-21] [WebCite Cache ID 6v9Bg9mMV]
10. Ventola CL. Social media and health care professionals: benefits, risks, and best practices. P T 2014 Jul;39(7):491-520 [FREE Full text] [Medline: 25083128]
11. Dizon DS, Graham D, Thompson MA, Johnson LJ, Johnston C, Fisch MJ, et al. Practical guidance: the use of social media in oncology practice. J Oncol Pract 2012 Sep;8(5):e114-e124 [FREE Full text] [doi: 10.1200/JOP.2012.000610] [Medline: 23277774]
12. Shcherbakova N, Shepherd M. Community pharmacists, Internet and social media: an empirical investigation. Res Social Adm Pharm 2014;10(6):e75-e85. [doi: 10.1016/j.sapharm.2013.11.007] [Medline: 24388002]
13. Thompson MA, Majhail NS, Wood WA, Perales M, Chaboissier M. Social media and the practicing hematologist: Twitter 101 for the busy healthcare provider. Curr Hematol Malig Rep 2015 Dec;10(4):405-412 [FREE Full text] [doi: 10.1007/s11899-015-0286-x] [Medline: 26449718]
14. Micieli R, Micieli JA. Twitter as a tool for ophthalmologists. Can J Ophthalmol 2012 Oct;47(5):410-413. [doi: 10.1016/j.jcjo.2012.05.005] [Medline: 23036540]
15. Desai T, Dhingra V, Shariff A, Shariff A, Lerma E, Singla P, et al. Quantifying the Twitter influence of third party commercial entities versus healthcare providers in thirteen medical conferences from 2011 - 2013. PLoS One 2016 Sep 26;11(9):e0162376 [FREE Full text] [doi: 10.1371/journal.pone.0162376] [Medline: 27668433]
16. Chung A, Woo H. Twitter in urology and other surgical specialties at global conferences. ANZ J Surg 2016 Apr;86(4):224-227. [doi: 10.1111/ans.13393] [Medline: 26631323]
17. Borgmann H, Woelm J, Merseburger A, Nestler T, Salem J, Brandt MP, et al. Qualitative Twitter analysis of participants, tweet strategies, and tweet content at a major urologic conference. Can Urol Assoc J 2016;10(1-2):39-44 [FREE Full text] [doi: 10.5489/cuaj.3322] [Medline: 26977205]
18. Momany M, Di Pietro A, Alexander WG, Barker BM, Harb OS, Kamoun S, et al. Meeting report: fungal genomics meets social media: highlights of the 28th fungal genetics conference at Asilomar. G3 (Bethesda) 2015 Dec;5(12):2523-2525 [FREE Full text] [Medline: 27034999]
19. Wilkinson SE, Basto MY, Perovic G, Lawrentschuk N, Murphy DG. The social media revolution is changing the conference experience: analytics and trends from eight international meetings. BJU Int 2015 May;115(5):839-846 [FREE Full text] [doi: 10.1111/bju.12910] [Medline: 25130687]
20. Awad NI, Cocchio C. Use of Twitter at a major national pharmacy conference. Am J Health Syst Pharm 2015 Jan 1;72(1):65-69. [doi: 10.2146/ajhp140175] [Medline: 25511841]
21. Mishori R, Levy B, Donvan B. Twitter use at a family medicine conference: analyzing #STFM13. Fam Med 2014 Sep;46(8):608-614 [FREE Full text] [Medline: 25163039]
22. McKendrick DR, Cumming GP, Lee AJ. Increased use of Twitter at a medical conference: a report and a review of the educational opportunities. J Med Internet Res 2012 Dec 11;14(6):e176 [FREE Full text] [doi: 10.2196/jmir.2144] [Medline: 23232765]
23. Tyrawski J, DeAndrea DC. Pharmaceutical companies and their drugs on social media: a content analysis of drug information on popular social media sites. J Med Internet Res 2015 Jun 1;17(6):e130 [FREE Full text] [doi: 10.2196/jmir.4357] [Medline: 26032738]
24. Jackson W, Park B, Toscani M, Hermes-DeSantis E. Analysis of social media interactions between pharmaceutical companies and consumers: the power of the like. Ther Innov Regul Sci 2015 May;49(3):387-391. [doi: 10.1177/2168479014561803] [Medline: 30222395]
25. Houston P. Pharmaceutical Executive. 2017 Aug 5. The Hear and Now for Pharma and Social Media URL: http://www.pharmexec.com/hear-and-now-pharma-and-social-media [accessed 2017-11-21] [WebCite Cache ID 6v8qwnKzA]
26. US Food and Drug Administration. 2014 Jan. Guidance for industry: fulfilling regulatory requirements for postmarketing submissions of interactive promotional media for prescription human and animal drugs and biologics [draft guidance] URL: https://www.fda.gov/downloads/drugs/guidances/ucm381352.pdf [accessed 2017-11-22] [WebCite Cache ID 6vAn4zShX]
27. US Food and Drug Administration. 2014 Jun. Guidance for industry: internet/social media platforms: correcting independent third-party misinformation about prescription drugs and medical devices [draft guidance] URL: https://www.fda.gov/downloads/drugs/guidances/ucm401079.pdf [accessed 2017-11-22] [WebCite Cache ID 6vAnOJaqy]

28. US Food and Drug Administration. 2014 Jun. Guidance for industry: internet/social media platforms with character space limitations - presenting risk and benefit information for prescription drugs and medical devices [draft guidance] URL: https://www.fda.gov/downloads/drugs/guidances/ucm401087.pdf [accessed 2017-11-22] [WebCite Cache ID 6vAncFz52]
29. US Food and Drug Administration. 2011 Dec. Guidance for industry: responding to unsolicited requests for off-label information about prescription drugs and medical devices [draft guidance] URL: https://www.fda.gov/downloads/drugs/guidances/ucm285145.pdf [accessed 2017-11-21] [WebCite Cache ID 6v8r42XSS]
30. Reisman M. Pharma twitters cautiously before advertising guidelines are released. P T 2010 Jul;35(7):373-416 [FREE Full text] [Medline: 20689623]
31. Bordoloi P, Gažo A, Paranjpe K, Clausen M, Fierro L. Medical information services in the age of social media and new customer channels. Drug Inf J 2011 Nov 1;45(6):811-818. [doi: 10.1177/009286151104500518]
32. Sermo. What is Sermo? URL: http://www.sermo.com/what-is-sermo/overview [accessed 2017-11-21] [WebCite Cache ID 6v9CdnhW5]

Abbreviations

FDA: Food and Drug Administration
HCP: health care professional
RUR: responding to unsolicited requests

Edited by G Eysenbach; submitted 12.12.17; peer-reviewed by S Ganesh, K Rolls; comments to author 20.03.18; revised version received 08.05.18; accepted 14.08.18; published 29.10.18.

Please cite as:
TenBarge AM, Riggins JL
Responding to Unsolicited Medical Requests from Health Care Professionals on Pharmaceutical Industry-Owned Social Media Sites: Three Pilot Studies
J Med Internet Res 2018;20(10):e285
URL: http://www.jmir.org/2018/10/e285/
doi:10.2196/jmir.9643
PMID:30373730

Original Paper

Group-Personalized Regression Models for Predicting Mental Health Scores From Objective Mobile Phone Data Streams: Observational Study

Niclas Palmius[1], MEng; Kate E A Saunders[2,3], MD, DPhil (Oxon); Oliver Carr[1], MEng; John R Geddes[2], MD; Guy M Goodwin[2,3], FMedSci; Maarten De Vos[1,4], MEng, PhD

[1] Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, United Kingdom
[2] Department of Psychiatry, University of Oxford, Oxford, United Kingdom
[3] Oxford Health National Health Service Foundation Trust, University of Oxford, Oxford, United Kingdom
[4] Sleep and Circadian Neuroscience Institute, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom

Corresponding Author:
Maarten De Vos, MEng, PhD
Institute of Biomedical Engineering
Department of Engineering Science
University of Oxford
Parks Road
Oxford, OX1 3PJ
United Kingdom
Phone: 44 1865617721
Fax: 44 1865617701
Email: maarten.devos@eng.ox.ac.uk

Abstract

Background: Objective behavioral markers of mental illness, often recorded through smartphones or wearable devices, have the potential to transform how mental health services are delivered and to help users monitor their own health. Linking objective markers to illness is commonly performed using population-level models, which assume that everyone is the same. The reality is that there are large levels of natural interindividual variability, both in terms of response to illness and in usual behavioral patterns, as well as intraindividual variability that these models do not consider.

Objective: The objective of this study was to demonstrate the utility of splitting the population into subsets of individuals that exhibit similar relationships between their objective markers and their mental states. Using these subsets, “group-personalized” models can be built for individuals based on other individuals to whom they are most similar.

Methods: We collected geolocation data from 59 participants who were part of the Automated Monitoring of Symptom Severity study at the University of Oxford. This was an observational data collection study. Participants were diagnosed with bipolar disorder (n=20); borderline personality disorder (n=17); or were healthy controls (n=22). Geolocation data were collected using a custom Android app installed on participants’ smartphones, and participants weekly reported their symptoms of depression using the 16-item quick inventory of depressive symptomatology questionnaire. Population-level models were built to estimate levels of depression using features derived from the geolocation data recorded from participants, and it was hypothesized that results could be improved by splitting individuals into subgroups with similar relationships between their behavioral features and depressive symptoms. We developed a new model using a Dirichlet process prior for splitting individuals into groups, with a Bayesian Lasso model in each group to link behavioral features with mental illness. The result is a model for each individual that incorporates information from other similar individuals to augment the limited training data available.

Results: The new group-personalized regression model showed a significant improvement over population-level models in predicting mental health severity (*P*<.001). Analysis of subgroups showed that different groups were characterized by different features derived from raw geolocation data.

Conclusions: This study demonstrates the importance of handling interindividual variability when developing models of mental illness. Population-level models do not capture nuances in how different individuals respond to illness, and the group-personalized model demonstrates a potential way to overcome these limitations when estimating mental state from objective behavioral features.

(J Med Internet Res 2018;20(10):e10194) doi:10.2196/10194

KEYWORDS

behavioral features; depression; geolocation; group-personalized model; interindividual variability; mental health; mental illness; objective behavioral markers

Introduction

One key research area in computing and mental health is finding relationships between objective markers of user behavior and mental state. Objective markers in this context may include physical activity levels, geographic movements, interaction with social networks, sleep quality and circadian regularity, and interaction with technology, among others. Commonly, these markers may be recorded continuously from smartphones, which provide easy access to an array of sensors that can provide such data in a completely passive way, although wearable devices and other sensors may also be used.

Objective behavioral markers have been explored in a number of previous studies. Physical activity level was one of the first behavioral markers to be widely studied. For example, Wielopolski et al [1] found significantly lower levels of physical activity in patients with acute unipolar depression than in healthy controls and that physical activity correlated with improvement in symptoms in 19 patients with depression. In contrast, Wang et al [2] found that the levels of activity in students were negatively correlated with self-rated loneliness scores, but not with self-ratings of depression. Use of technology, especially mobile phones, has also been studied, with Saeb et al [3] reporting that both frequency of phone usage and total time spent interacting with the phone correlated with levels of depression in a community cohort of 21 individuals, with more depressed individuals interacting more with their phones. Mehrotra et al [4] also reported that the way users interacted with their phones, such as the number of notifications responded to and the time taken to respond, strongly correlated with depression scores from 25 individuals. Another promising behavioral data source is the geographic movement of individuals, which indicates activity on a higher level with potentially greater accuracy than physical activity recorded through accelerometry. While people may not carry their phones during all physical activity, they are likely to carry them when moving across larger geographic distances. Among the earliest work on using geolocation for mental state estimation, Grünerbl et al [5-6] demonstrated that it is possible to use geolocation-derived features to detect episodes in patients with bipolar disorder. Saeb et al [3] further showed that features derived from geolocation data correlated with depression levels in individuals recruited over the internet; the results were later replicated with a sample of students [7].

The previous work summarized above has focused mainly on finding population-level models [3,6,7] or correlations [2,4] that link measured behavioral markers to mental state. Specifically, most studies either attempt to classify groups of patients by condition or perform regression to estimate the patient's mental state (usually using patient self-ratings—a limitation explored further in the Discussion). Population-level models use all available data from a given population to link the mental states of individuals to their behavioral features. This is also the case in the popular approach of classifying patient groups based on their behavioral symptoms. It is implicitly assumed that the same set of features will have discriminative power across the whole population. While this is a promising approach, it is widely accepted that natural variability in the usual behavioral patterns of different individuals, or differences in their behavior for a given level of illness, are major limitations of these models. Mohr et al referred to this as the "curse of variability" [8]. A recent review by Berrouiguet et al [9] also highlighted the need to move toward a personalized approach for developing digital tools, especially in mental healthcare. While agreeing that different individuals may have different models that link their behavior to their mental states, how best to define such individual models is still an open question. Training "fully personalized" models on any available training data for each individual (for example, by asking the individual to provide objective data and questionnaire responses for a number of weeks) would in theory produce the most accurate model for that individual because it would eliminate any interindividual variability. Grünerbl et al [5] used this approach to identify mental state in patients with bipolar disorder from geolocation-derived features. Similarly, Canzian and Musolesi [10] reported correlations calculated for individual participants between geolocation-derived features and daily self-reported depression scores in 28 participants who provided at least 20 usable data points. Significant correlations were shown between depression scores and the maximum distance between any 2 locations over the last 14 days for 18 of the 28 participants; this again demonstrates the utility of geolocation as a predictor for depression. However, in practice, this approach is limited by the amount of data that would be required from each individual to train the model, may not generalize to unseen states, and will likely overfit available data, leading to reduced out-of-sample performance.

A compromise between population-level models and fully personalized models is to create groups of individuals with similar characteristics [11]. New individuals can then be allocated to an existing group based on their similarity to individuals in that group. If limited training data are available from new individuals, then their data can be combined with data from similar individuals. In a large study of over 18,000 people, Servia-Rodríguez et al [12] found correlations between most of the available demographics (age, gender, occupation, etc) and smartphone-recorded behavioral data. This means that another way of achieving personalized models is by conditioning on available demographics, but this approach is limited by the level of detail in the recorded demographic data. Hong et al have also previously demonstrated that demographic features poorly predict sensor data similarity [13]. Likely more subtle subgroups exist in the population, with similar correlations between behavioral data and mental state, but these groups may not be identifiable by explicit demographic features alone. The challenge is finding these subgroups, which can be thought of as “behavioral phenotypes” within the population. In addition to providing improved models of patient health, understanding the characteristics of these phenotypes may help develop our understanding of how to classify or subclassify mental illness [14].

Personalization by finding similar individuals in the population has long been considered in activity recognition from sensor data [11]. For example, Lane et al [15] developed, and Abdullah et al [16] extended, an approach for finding groups of individuals based on the similarity of their demographics, lifestyle, and sensor data. An advantage of this method is that it does not require any calibration data for new individuals. However, clustering on distributions of sensor data and not on relationships between sensor data and model output (in the present case, mental state) may miss important differences or similarities among individuals.

This paper proposes a novel, data-driven approach toward group-personalized regression. We describe an extension to classical linear regression whereby groups of individuals who exhibit similar linear models linking objective behavioral data and mental state are automatically identified. This model’s ability to identify groups of similar individuals is demonstrated by comparing traditional population-level regression with the novel approach for the estimation of self-reported levels of depression from objective geolocation-derived features.

Methods

Data Collection

Data were collected from participants in the Automated Monitoring of Symptom Severity (AMoSS) study at the University of Oxford [17,18]. The AMoSS study was approved by the Research Ethics Committee of the East of England (reference 13/EE/0288), and all participants provided written informed consent. During the AMoSS study, a range of behavioral data were collected from patients diagnosed with bipolar disorder and borderline personality disorder as well as healthy control individuals without any symptoms of mental disorder. All participants were screened by an experienced psychiatrist using the Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders, Fourth edition. Objective behavioral data were collected from a custom Android-based app as well as other wearable devices. The app recorded individuals’ activity levels, geographic movements (described below), light exposure, and social interaction. We also collected physiological data during some parts of the study, as presented by Carr et al [19,20]. More details about the study design are available in our previous work [17].

Mental State Reporting

Participants self-reported their mental state throughout their participation in the study, using a variety of clinically validated questionnaires, administered weekly. One questionnaire, the 16-item Quick Inventory of Depressive Symptomatology (QIDS) Self-Report questionnaire [21], was used to assess the level of depression in individuals. This questionnaire asks 16 questions assessing depression based on clinical diagnostic criteria and provides a single score from 0 (no symptoms of depression) to 27 (severe symptoms of depression). A score over 10 is considered a suitable threshold for clinically significant depression. A simple custom mood questionnaire was also administered daily on the app, which Tsanas et al have shown to correlate well with QIDS [22,23].

Location Data Overview

One of the key sources of behavioral data that can be easily recorded through a mobile phone is the individual’s geographic movements. We described in detail the collection, noise-removal, processing, and feature extraction from geolocation data in our previous work [18], wherein we demonstrated that several features can clearly discriminate between nondepressed and depressed weeks in patients with bipolar disorder.

Features were extracted from preprocessed geolocation data. In total, 10 features described in Table 1 were extracted and used in this paper, with full details available in our previous work [18]. All features were calculated on full calendar weeks of data (Monday to Sunday).

Data Inclusion

For the analysis presented here, it is important to have multiple data points available for each individual. For this reason, only participants who provided ≥6 weeks of geolocation data with associated QIDS scores were included in the analysis. A total of 59 participants provided the required minimum of 6 labeled weeks of data. Demographic characteristics of the included participants are shown in Table 2, which also shows summary statistics of the data available for analysis. Healthy control participants had the lowest mean QIDS scores and least variability. Participants with borderline personality disorder had the highest mean QIDS scores and also the highest variability, with participants with bipolar disorder between the two.

Standard Population-Level Regression Model

Previous studies estimating continuous mental health severity from objective markers [3], commonly work with standard linear regression models [24] of the form: $y=\beta x^T+\mu$.

In our case, x would be a vector of the geolocation-derived features in Table 1, and *y* is the predicted QIDS score. This regression model therefore forms the core baseline model for comparison of the group-personalized model described in the following section. Linear models as given above are prone to overfitting the available training data resulting in poor out-of-sample performance. One well-known method to reduce the overfitting of the model is the Lasso by Tibshirani [25], which adds an ℓ_1 regularization term to select only the most important features to include in the model. In its Lagrangian dual form [26], the Lasso is parameterized by λ. When λ is large, many (or all) of the coefficients of β will be reduced to exactly zero.

Table 1. Summary of features extracted from preprocessed geolocation data.

Feature name	Description
Location variance	A measure of the variance in the location coordinates visited.
Number of clusters	The number of unique locations visited.
Entropy of locations	The information-theoretic entropy calculated on the proportion of time spent in each of the locations visited.
Normalized entropy	The entropy of locations feature normalized by dividing by the log of the number of location visited, resulting in a feature ranging between 0 and 1, which is less correlated with the number of clusters feature.
Home stay	The percentage of time that the individual is recorded at home.
Transition time	The percentage of time that the individual is recorded traveling between locations.
Total distance	The total distance traveled by the individual.
Diurnal movement	A measure of the diurnal regularity in the movements of the individual, calculated from the power in sinusoids fitted to the data with periods around 24 hours.
Diurnal movement on normalized coordinates	Similar to the diurnal movement feature, but calculated on normalized coordinates, making it less sensitive to the different distances that individuals may travel.
Diurnal movement on distance from home	Similar to the diurnal movement and diurnal movement on normalized coordinates features, but calculated on the single dimensional distance of the current location coordinates from the home location of the individual.

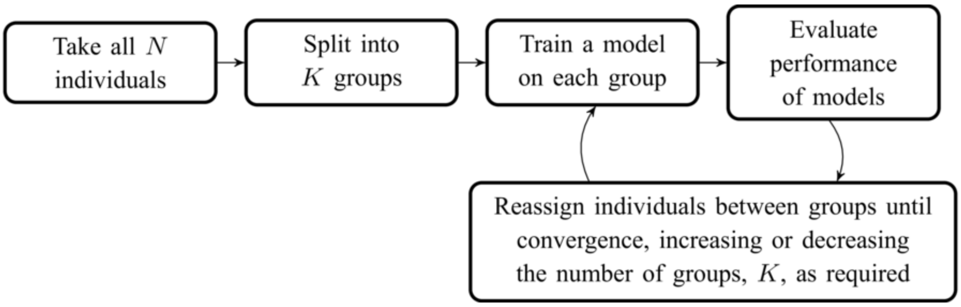
Table 2. Demographic and data characteristics of participants included in the analysis.

Characteristic	Healthy controls (n=22)	Bipolar disorder patients (n=20)	Borderline personality disorder patients (n=17)	Total (n=59)
Gender, n				
Male	7	7	1	15
Female	15	13	16	44
Age, median (IQR[a])	42 (12)	44 (20)	38 (9.75)	41 (15.75)
Body mass index, median (IQR)	24 (5.37)	27 (4.22)	31 (10.25)	26 (8.50)
Weeks of data per participant, median (IQR)	17 (15.0)	16 (19.5)	20 (24.5)	19 (19.75)
QIDS[b] mean, median (IQR)	2 (1.92)	5 (6.74)	14 (4.67)	4 (9.96)
QIDS range, median (IQR)	3 (2.07)	7 (6.64)	10 (7.34)	5 (7.52)

[a] IQR: interquartile range.

[b] QIDS: Quick Inventory of Depressive Symptomatology.

Figure 1. High-level overview of the group-personalized model.



Group-Personalized Regression Model

Here, we will introduce a method for splitting the total population into groups of individuals that have similar relationships between their geolocation-derived features and their QIDS scores. Each group will be represented by a Lasso model where the regression coefficients β and offset μ will differ between subsets. This group-personalized regression model has been developed to improve regression performance by finding models in subgroups that fit individuals better than the population-level model trained across all available individuals, while not training an individual model for each subject.

Figure 1 depicts the high-level framework of the group-personalized model used to find groups of individuals that have similar regression models linking mental states and behavioral features. During initialization, all available individuals are randomly split into groups, and a Lasso regularized linear regression model is trained for each group. The performance of all individuals under the model for each group is then evaluated, and individuals are reassigned between groups as appropriate. Note that the number of groups is dynamic and optimized as the model is run. This process is repeated to convergence. Using Lasso regularized linear regression as the core of the group-personalized model allows identification of the most relevant subset of features for each group.

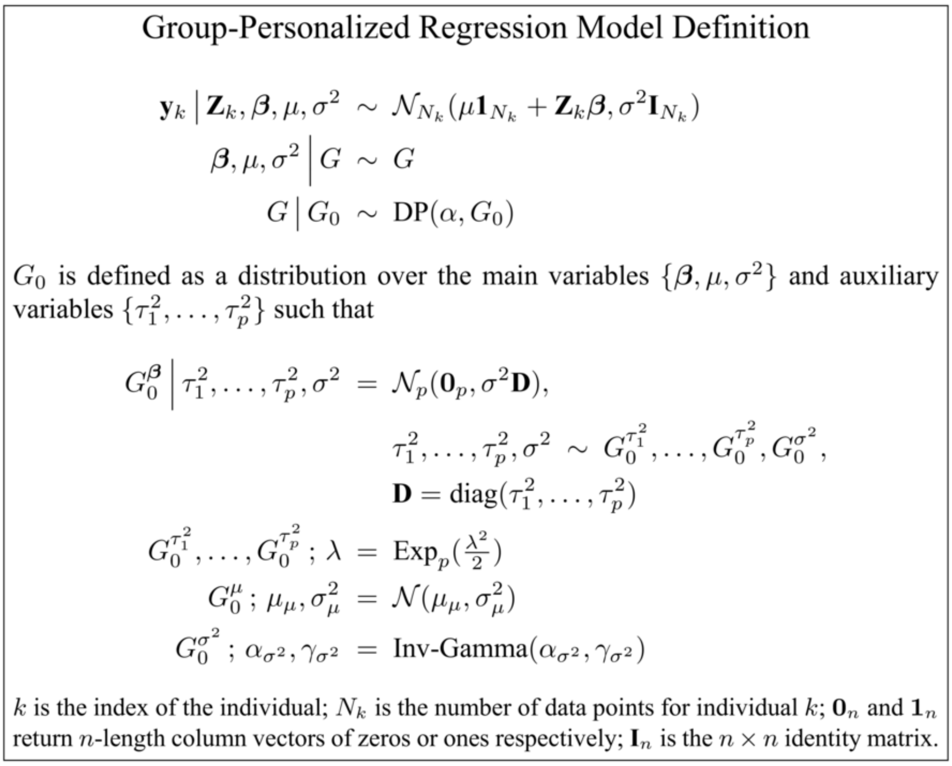
In a practical realization of the model depicted in Figure 1, a framework that combines the predictive model and clustering is required. The Dirichlet process (DP) provides a suitable framework [27]. A DP mixture model splits individuals into distinct groups, trains a model on individuals in each group, and then allows individuals to switch groups until clustering is optimized. One feature of the DP is that the number of clusters does not need to be specified in advance, and individuals may create new clusters if they do not fit well into existing ones.

The DP is a Bayesian model where any well-specified generative model can be used as the model for individual clusters (with the requirement that model priors can be sampled and that the joint likelihood can be calculated). Park and Casella [28] have previously presented a Bayesian version of the Lasso. In this work, we extend the Bayesian Lasso to operate over multiple individuals and apply a DP prior to provide clustering of individuals based on the relationship between their behavioral features and their levels of self-reported depression.

The full generative form of the group-personalized regression model is given in Figure 2. Free parameters in the model are the α concentration parameter of the DP; the λ regularization parameter of the Lasso model for each group; the μ_μ and σ_μ^2 priors of the μ variable (the regression offset); and the α_σ^2 and γ_σ^2 priors of the σ^2 variable (the noise in the model). The α concentration parameter and the λ regularization parameter were both optimized by a grid search maximizing the likelihood of the model's joint probability. Parameter values of λ=0.4 and α=0.0001 were found optimal and are used to generate the results presented here. Other free priors have much less impact on the model and were set to $\mu_\mu=0$; $\sigma_\mu^2=10$; $\alpha_\sigma^2=1$; and $\gamma_\sigma^2=1$.

The group-personalized regression model defined in Figure 2 can be sampled using a standard implementation of Algorithm 8 by Neal [27,29]. In each iteration of sampling the DP, individual grouped model variables are Gibbs sampled from the distributions given in Figure 3.

Figure 2. Formal definition of the group-personalized regression model.



Group-Personalized Regression Model Definition

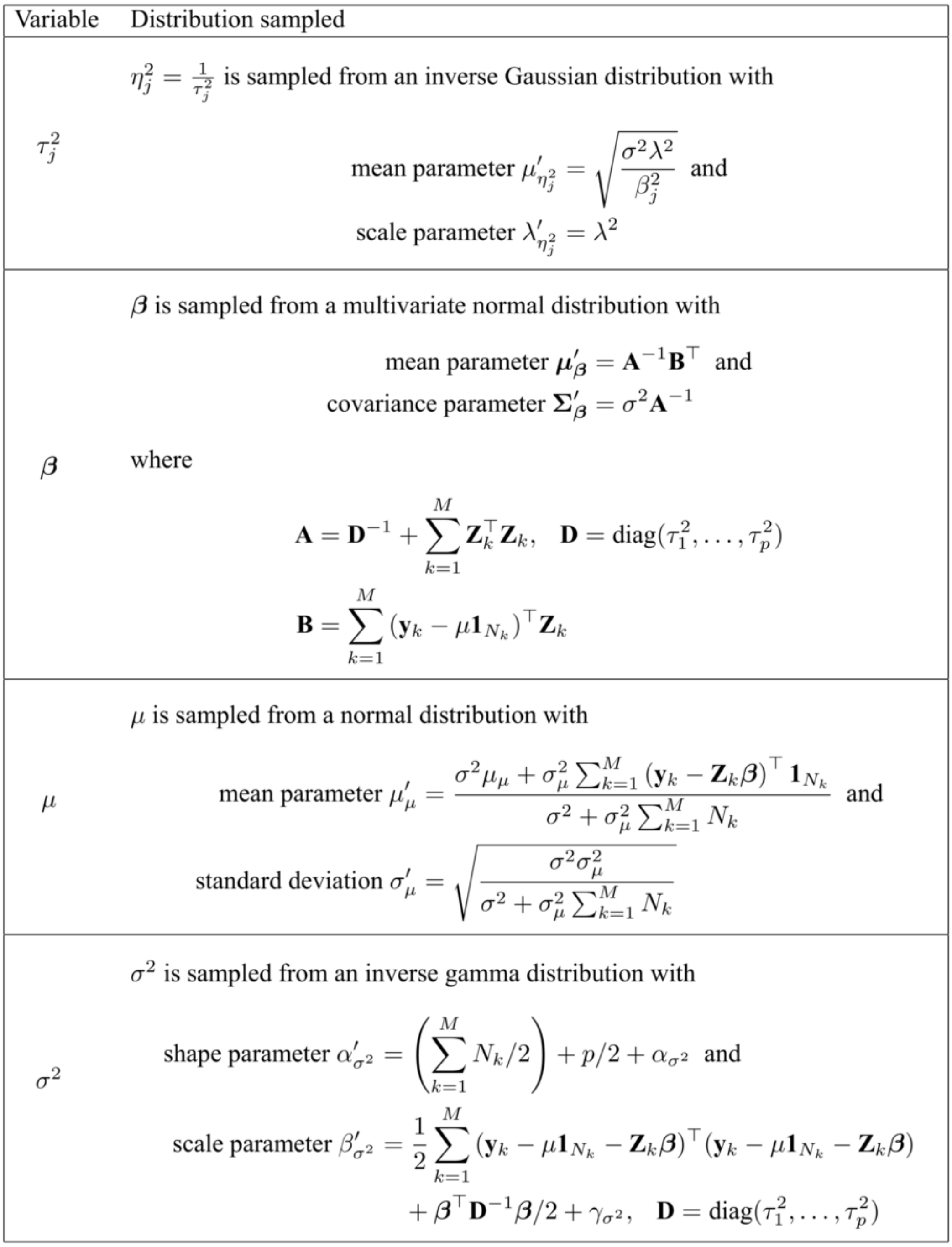
$$
\begin{aligned}
\mathbf{y}_k \,\big|\, \mathbf{Z}_k, \boldsymbol{\beta}, \mu, \sigma^2 &\sim \mathcal{N}_{N_k}(\mu \mathbf{1}_{N_k} + \mathbf{Z}_k \boldsymbol{\beta}, \sigma^2 \mathbf{I}_{N_k}) \\
\boldsymbol{\beta}, \mu, \sigma^2 \,\big|\, G &\sim G \\
G \,\big|\, G_0 &\sim \mathrm{DP}(\alpha, G_0)
\end{aligned}
$$

G_0 is defined as a distribution over the main variables $\{\boldsymbol{\beta}, \mu, \sigma^2\}$ and auxiliary variables $\{\tau_1^2, \ldots, \tau_p^2\}$ such that

$$
\begin{aligned}
G_0^{\boldsymbol{\beta}} \,\big|\, \tau_1^2, \ldots, \tau_p^2, \sigma^2 &= \mathcal{N}_p(\mathbf{0}_p, \sigma^2 \mathbf{D}), \\
&\quad \tau_1^2, \ldots, \tau_p^2, \sigma^2 \sim G_0^{\tau_1^2}, \ldots, G_0^{\tau_p^2}, G_0^{\sigma^2}, \\
&\quad \mathbf{D} = \mathrm{diag}(\tau_1^2, \ldots, \tau_p^2) \\
G_0^{\tau_1^2}, \ldots, G_0^{\tau_p^2} ; \lambda &= \mathrm{Exp}_p(\tfrac{\lambda^2}{2}) \\
G_0^{\mu} ; \mu_\mu, \sigma_\mu^2 &= \mathcal{N}(\mu_\mu, \sigma_\mu^2) \\
G_0^{\sigma^2} ; \alpha_{\sigma^2}, \gamma_{\sigma^2} &= \text{Inv-Gamma}(\alpha_{\sigma^2}, \gamma_{\sigma^2})
\end{aligned}
$$

k is the index of the individual; N_k is the number of data points for individual k; $\mathbf{0}_n$ and $\mathbf{1}_n$ return n-length column vectors of zeros or ones respectively; $\mathbf{I}_n$ is the $n \times n$ identity matrix.

Figure 3. Gibbs sampler distributions for variables in the group-personalized regression model.



Variable	Distribution sampled
τ_j^2	$\eta_j^2 = \frac{1}{\tau_j^2}$ is sampled from an inverse Gaussian distribution with mean parameter $\mu'_{\eta_j^2} = \sqrt{\frac{\sigma^2\lambda^2}{\beta_j^2}}$ and scale parameter $\lambda'_{\eta_j^2} = \lambda^2$
$\boldsymbol{\beta}$	$\boldsymbol{\beta}$ is sampled from a multivariate normal distribution with mean parameter $\boldsymbol{\mu}'_{\boldsymbol{\beta}} = \mathbf{A}^{-1}\mathbf{B}^\top$ and covariance parameter $\boldsymbol{\Sigma}'_{\boldsymbol{\beta}} = \sigma^2\mathbf{A}^{-1}$ where $\mathbf{A} = \mathbf{D}^{-1} + \sum_{k=1}^{M} \mathbf{Z}_k^\top \mathbf{Z}_k, \quad \mathbf{D} = \mathrm{diag}(\tau_1^2, \ldots, \tau_p^2)$ $\mathbf{B} = \sum_{k=1}^{M} (\mathbf{y}_k - \mu\mathbf{1}_{N_k})^\top \mathbf{Z}_k$
μ	μ is sampled from a normal distribution with mean parameter $\mu'_\mu = \frac{\sigma^2\mu_\mu + \sigma_\mu^2 \sum_{k=1}^{M} (\mathbf{y}_k - \mathbf{Z}_k\boldsymbol{\beta})^\top \mathbf{1}_{N_k}}{\sigma^2 + \sigma_\mu^2 \sum_{k=1}^{M} N_k}$ and standard deviation $\sigma'_\mu = \sqrt{\frac{\sigma^2\sigma_\mu^2}{\sigma^2 + \sigma_\mu^2 \sum_{k=1}^{M} N_k}}$
σ^2	σ^2 is sampled from an inverse gamma distribution with shape parameter $\alpha'_{\sigma^2} = \left(\sum_{k=1}^{M} N_k/2\right) + p/2 + \alpha_{\sigma^2}$ and scale parameter $\beta'_{\sigma^2} = \frac{1}{2}\sum_{k=1}^{M} (\mathbf{y}_k - \mu\mathbf{1}_{N_k} - \mathbf{Z}_k\boldsymbol{\beta})^\top (\mathbf{y}_k - \mu\mathbf{1}_{N_k} - \mathbf{Z}_k\boldsymbol{\beta})$ $+ \boldsymbol{\beta}^\top \mathbf{D}^{-1}\boldsymbol{\beta}/2 + \gamma_{\sigma^2}, \quad \mathbf{D} = \mathrm{diag}(\tau_1^2, \ldots, \tau_p^2)$

Model Evaluation

Evaluation Framework

This study's aim was to predict QIDS scores from geolocation-derived features and to compare the performance of population-level, fully personalized, and group-personalized regression models for this purpose. For this reason, several models described below have been implemented to present comparative results.

Most models described in the following sections were tested in a "leave-one-participant-out" framework where each participant is left out and the relevant model is trained on data from other relevant participants (as described for each model in turn).

In practice, fully personalized and group-personalized models generally require a certain level of calibration data (behavioral features and QIDS scores). To simulate this, in most models, a subset of data from the start of the recording for the left-out test individual were included in the model's training (but were excluded when evaluating test performance). The number of weeks of calibration data included in the model training is defined as the first half of the available data for the test individual up to a maximum of 8 weeks. For example, for an individual who provided 10 weeks of data, the first 5 weeks are used as calibration data included in model training and the last 5 weeks are used for evaluation. For an individual who provided ≥16 weeks of data, the first 8 weeks are used as calibration data and all remaining data are used for evaluation.

In all cases except the clustering model, standard Bayesian Lasso models by Park and Casella [28] were trained on the calibration data for the left out individual and all data from other relevant individuals. The Bayesian Lasso model was trained with regularization parameter λ=0.4 (the same value found optimal for running the group-personalized model) for 1000 iterations. Because the Bayesian Lasso is implemented using a Gibbs sampler, this leads to 1000 samples of model coefficients. The model's performance was evaluated in each iteration by applying the Bayesian Lasso with the sampled coefficients on the evaluation data for the left out individual. The mean absolute error (MAE) of the estimated QIDS scores and self-reported values was calculated in each iteration, and overall performance was evaluated as the mean MAE over all iterations. Improvements in performance were compared using a single-tailed paired *t* test of the mean MAE over all iterations for each individual.

Population-Level Model

The population-level model was tested by training a Bayesian Lasso model on the calibration data from the test individual and all data available from all other individuals.

Group-Personalized Model

The group-personalized model was introduced by first finding groups of individuals in the population that have similar models linking their geolocation-derived features with their QIDS scores. To find these groups, the group-personalized regression model was run using Neal's Algorithm 8 for 5000 iterations with all data available from all individuals.

To evaluate the performance of the optimal groups found using the group-personalized model, a Bayesian Lasso model was trained on the calibration data from the test individual and all data available from other individuals in the same group.

Note that if an individual ends up in a group containing just himself or herself, then the group-personalized model will be trained just on the calibration data from that individual, so that it reduces to the fully personalized model described below.

Group-Personalized Model with Clusters Allocated Using Calibration Data

The group-personalized model tested retrospectively, as described above, is useful to demonstrate principles of the model's operation and the clinical relevance of the groups found.

To apply the model prospectively, only calibration data from new individuals must be used to allocate them into one of the groups. To test this, each individual was left out in turn, and separate Bayesian Lasso models were trained for each group. For the group that the left-out individual was originally assigned to, the Bayesian Lasso model was trained on that group's remaining individuals (if any).

QIDS scores were estimated from the calibration data for the left-out individual using the model trained on each of the groups. The mean MAE of the estimated QIDS scores using the model for each group was evaluated, and the individual was allocated to the group that provided optimal performance.

To provide the final prediction, a new Bayesian Lasso model was trained using the calibration data from the left-out individual and all data from other individuals assigned to the allocated group.

Fully Personalized Model Using All Available Data

The literature [5,6,30] commonly uses cross-validation over all available data points from an individual to demonstrate a "personalized" model. This was implemented using random subsampling of data from each individual, where a Bayesian Lasso model was trained using 80% of data randomly selected from the test individual, with results presented on estimation of the remaining 20%. This split was repeated 10 times.

Fully Personalized Model Using Calibration Data Only

While the fully personalized model using all available data is commonly presented in the literature, it may not provide a fair representation of the model's accuracy in practice because it does not demonstrate how well the model will generalize when trained with limited calibration data.

As an alternative, the fully personalized model using only calibration data was tested by training a Bayesian Lasso model on just the calibration data from the test individual.

Clustering Model

A comparative model was tested using the clustering method described by Lane et al [15] and Abdullah et al [16]. To provide a comparative result, only the part of the method that clusters individuals by similarity of their feature values was included. The clustering method by Lane et al and Abdullah et al used a locality-sensitive hashing method known as random projection [31] as a similarity measure of feature values between all pairs of individuals. The random projection method samples random values in the feature space and calculates the distance from these random values to extracted feature values for each individual. By repeating this process multiple times, features from similar individuals will commonly be closest to the same randomly sampled values. Lane et al and Abdullah et al used the resulting similarity matrix to condition an online boosting classification algorithm. Because the group-personalized model concerns regression rather than classification, this was replaced with a regression equivalent [32].

Because the clustering model does not use the relationship between behavioral features and model output, all available data from all individuals, including the test individual, were included when assessing similarity.

Results

Extracted Feature Properties

Raw feature values calculated from the geolocation data for included participants are shown in Figure 4 for 3 selected features (the number of locations visited on the left; entropy in the middle; and the percentage of time recorded at home on the right); highlighted are 6 participants (2 from each cohort). Standard population-level linear regression models predicting the QIDS score from all available data for each of the features individually are shown overlaid. The general trends follow what is expected from the literature and has been presented in previous work. More depressed individuals tend to visit fewer locations and stay at home more. As a measure of variability in the locations visited, entropy tends to be lower in more depressed individuals, indicating that depressed individuals have less regularity in their routines.

Clearly, however, these trends are weak, and the features contain a high level of noise. Models trained over all available data (such as the population-level linear models shown in Figure 4) would therefore tend to estimate depression poorly on behavioral features.

The 6 participants highlighted in Figure 4 indicate that data from specific individuals tended to occur in clusters within the whole dataset. In some cases, such as the green borderline personality disorder individual, personalizing the model using the number of clusters visited should far outperform the population-level model.

Model Evaluation

The group-personalized regression model was run over all individuals as described above. In total, 17 groups were found, each containing between 1 and 9 individuals. Figure 5 shows how individuals were allocated to these groups. Each group is shown as a row, with the shading of the markers indicating how many individuals are from each cohort in that group. Groups 1-4 are predominantly healthy control individuals, together with participants with bipolar disorder who exhibit low variability in their QIDS scores. Groups 5 and 6 are predominantly individuals with bipolar disorder, and groups 7-10 are predominantly individuals with borderline personality disorder, all displaying greater variability in their QIDS scores. To groups 11-17, 9 individuals were assigned, with only 1 or 2 individuals in each group. These individuals have been shown as unassigned because they did not fit well into any of the other groups, and therefore solid conclusions cannot be drawn. As more individuals become available for analysis, these individuals would likely be assigned to larger groups.

The number of groups found (17 for 59 participants) indicates the high level of variability in the relationships between behavioral data and mental state for individuals in the study. The number of groups may also be affected by properties of the DP prior on smaller sample sizes. Asymptotically, the DP exhibits a “rich-gets-richer” property, where larger clusters tend to become larger [33]. This also has the side effect that with more data points, the number of clusters relative to the number of samples decreases (more precisely, the expected number of clusters grows logarithmically with the number of data points, proportional to the concentration parameter in the DP prior). With fewer data points, none of the clusters are yet large enough to sufficiently attract data points. This means that as more data become available to train the model, the relative number of clusters should decrease, thus improving the model’s stability.

These findings imply some overlap between healthy control individuals and individuals with bipolar disorder. This fits with the informal observation that individuals with bipolar disorder exhibit normal mood when well. By contrast, there is very little overlap between healthy control and borderline personality disorder group membership. This finding broadly aligns with other subjective measures of mental state, sampled much more frequently than weekly, which show a gradient of abnormal mood where healthy control<bipolar disorder<borderline personality disorder [19,20]. How more frequent sampling would combine with geolocation data is an important future question.

One of the key advantages of finding similar groups of individuals within the population is that models for each group found may indicate different characteristics of patient subgroups. This can be explored by inspecting the coefficient values sampled for each group. For this, a Bayesian Lasso model was run on all individuals allocated to each group. This results in the sampled coefficient values shown in Figure 6 for groups 7 and 9 in Figure 5.

The group on the left in Figure 6 is characterized mainly by the diurnal movement, transition time, and total distance features, while the group on the right is characterized by the number of clusters visited. In both cases, the other features are effectively removed from the model. Pearson correlation coefficients of each feature with the reported QIDS scores for individuals in the group are shown on the right of each feature. Statistically significant correlation coefficients (*P*<.01) are shown in bold. This shows that the group-personalized model tends to pick out the most correlated features for each group, although this may not necessarily be the case as two features with low correlations may be predictive when combined in a regression model.

Figure 4. All available data for 3 of the geolocation-derived features, with the data from 6 individuals highlighted, 2 from each cohort in the study. (HC: healthy control, BD: bipolar disorder, BPD: borderline personality disorder, QIDS: Quick Inventory of Depressive Symptomatology).

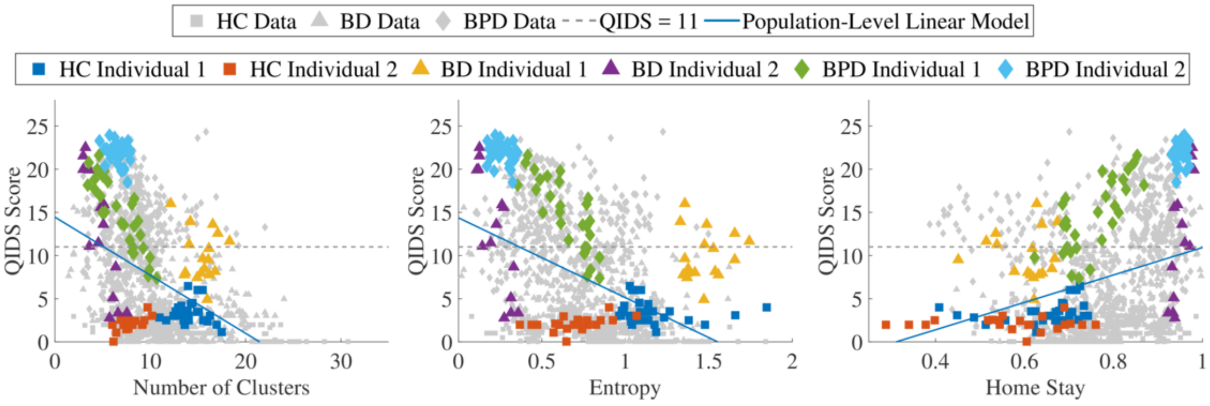


Figure 5. Allocations of individuals to different groups, showing the cohort of each individual. (HC: healthy control, BD: bipolar disorder, BPD: borderline personality disorder).

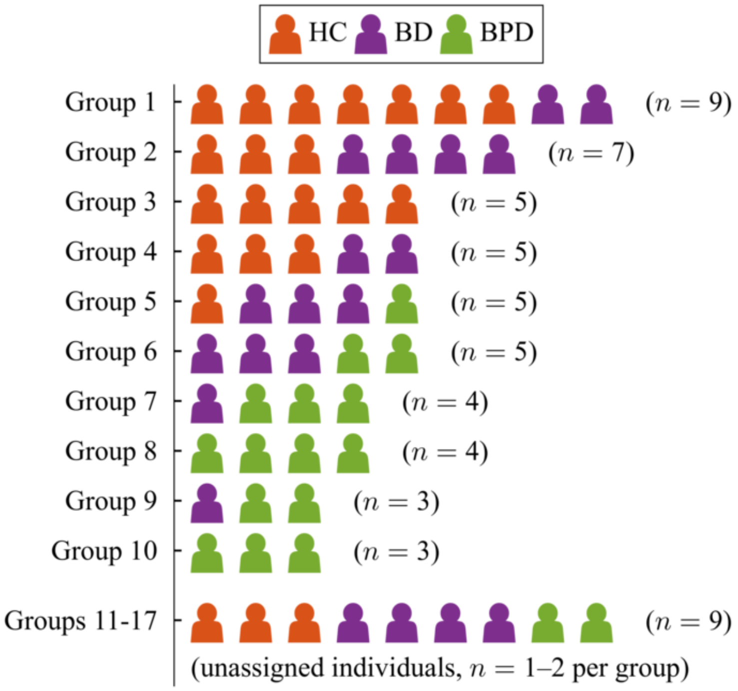
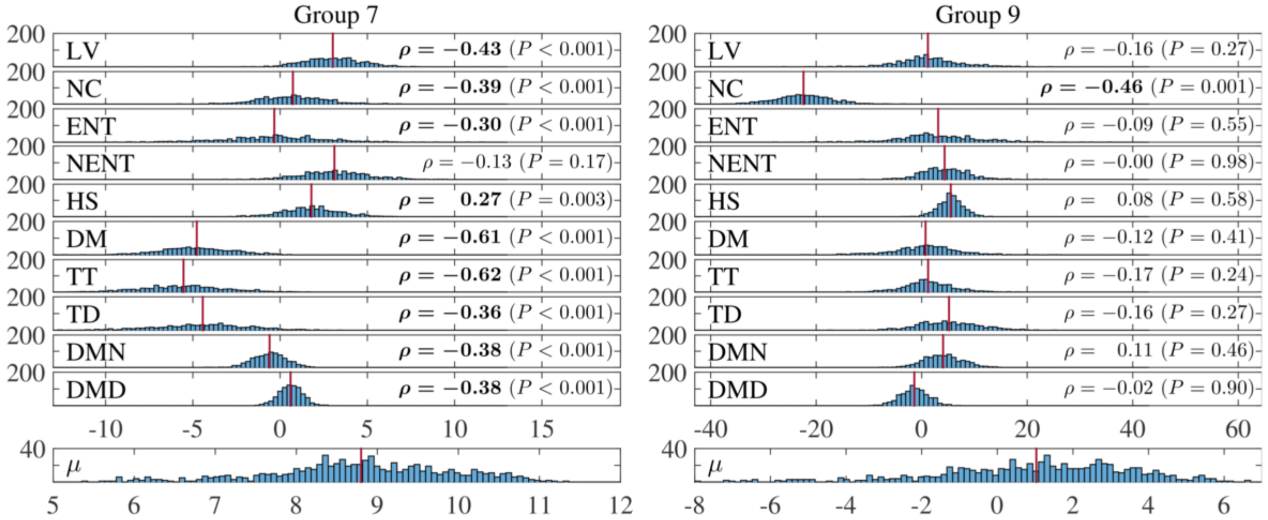


Figure 6. Sampled coefficient values for features in 2 groups found by the group-personalized model. (LV: location variance; NC: number of clusters; ENT: entropy of locations; NENT: normalized entropy; HS: home stay; DM: diurnal movement; TT: transition time; TD: total distance; DMN: diurnal movement on normalized coordinates; DMD: diurnal movement on distance from home).



Inspection of the model coefficients in each of the clusters found indicate that they all model different characteristics of the range of relationships between behavioral data and QIDS scores. Most groups are also characterized by ≤3 main features, indicating the interpretability of the models found.

Performance of the different models tested across all individuals is summarized in Table 3. Results are presented as MAE, mean (SD) for all models. Overall estimation accuracy is increased from the population-level model by using both fully personalized and group-personalized models. Optimal results are achieved with the fully personalized model, trained using subsamples of all available data. This is not surprising since the full range of values is likely to be included in the model's training. The group-personalized model with individuals allocated to their optimal groups improves over the population-level model and is similar to the fully personalized model. A single-tailed paired *t* test confirms that the performance improvement using the group-personalized model with optimized clusters from the population-level model is significant (P<.001). The group-personalized model with individuals assigned to groups using only their calibration data also performs significantly better than the population-level model. It also improves slightly on the fully personalized model trained only on the same calibration data, but the improvement is only mildly significant. Performance of the fully personalized model trained using only calibration data indicates the difficulty of generalizability of a model trained using limited data. The clustering method by Lane et al [15] and Abdullah et al [16] improves on the population-level model for bipolar disorder and borderline personality disorder participants, but does not perform better overall. This is can be explained by the method's entire basis on similarity of input features, not on the similarity of the relationship between the input features and QIDS scores.

Figure 7 shows the distributions of performance under 3 of the models tested: the top row shows the MAE of estimations made using the population-level model; the second row shows the MAE of estimations made using the fully personalized model; and the bottom row shows the MAE of estimations made using the group-personalized model. In all 3 graphs, bars are shaded in proportions corresponding to the cohort of individuals in that bar. Individuals with borderline personality disorder tend to perform worst under the population-level model.

Table 3. Mean absolute error of Quick Inventory of Depressive Symptomatology score estimation.

Model	HC[a], mean (SD)	BD[b] patients, mean (SD)	BPD[c] patients, mean (SD)	Overall, mean (SD)	Significance of reduction in overall mean absolute error compared to reference models	
					Population-level model (*P* value)	Fully personalized model trained on calibration data (*P* value)
Population-level model	4.86 (2.54)	4.74 (2.07)	6.43 (2.58)	5.27 (2.48)	—[d]	—
Fully personalized model using cross-validation validation over all data points	0.80 (0.76)	2.27 (1.44)	3.05 (1.67)	1.94 (1.60)	<.001	<.001
Fully personalized model trained on calibration data	1.06 (0.75)	3.67 (2.60)	4.38 (2.43)	2.90 (2.49)	<.001	—
Clustering based on Lane et al [15] and Abdullah et al [16]	5.33 (3.11)	4.50 (2.30)	6.15 (2.88)	5.29 (2.82)	—	—
Group-personalized model with optimized clusters	0.83 (0.52)	2.30 (1.96)	2.82 (1.28)	1.90 (1.60)	<.001	<.001
Group-personalized model with clusters allocated using calibration data	0.86 (0.46)	3.30 (3.09)	3.75 (2.07)	2.52 (2.47)	<.001	.02

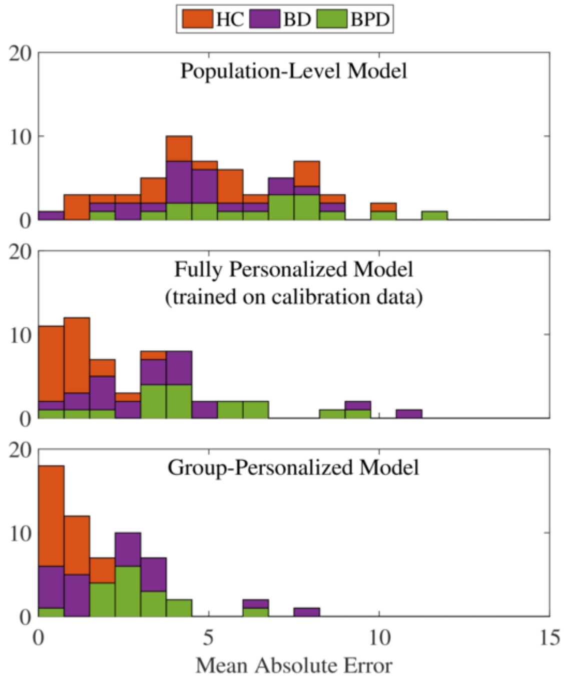
[a] HC: healthy control.

[b] BD: bipolar disorder.

[c] BPD: borderline personality disorder.

[d] Not applicable.

Figure 7. Mean absolute error of Quick Inventory of Depressive Symptomatology score estimation using 3 of the models in Table 3. (HC: healthy control; BD: bipolar disorder; BPD: borderline personality disorder).



The best performance under the group-personalized model is from healthy control participants, who showed quite variable performance under the population-level model. However, this should be viewed with caution because healthy control participants also tended to show very low QIDS scores with very little variation, as demonstrated in Table 2, meaning that a model may just estimate a constant low value. Similarly, some participants with bipolar disorder also perform very well under the group-personalized model, but as shown in Figure 5, some individuals with bipolar disorder were allocated with healthy control individuals, indicating that they may also have relatively constant low QIDS scores. The remaining individuals with bipolar disorder and borderline personality disorder clearly showed great improvement under the group-personalized model.

Discussion

Principal Findings

Our previous work demonstrated the utility of using geolocation-derived features to classify weeks of depression in participants with bipolar disorder. In this paper, we investigated to what extent those features can be used as proxy for the level of depression in subjects with bipolar disorder and borderline personality disorder. We have shown that using a population-based estimate is suboptimal because variability in behavioral patterns between subjects is too high, as shown in Figure 4. While fully personalizing a model for each individual might make sense, doing so requires too much data to be practicable in a prospective fashion and provides little clinical insight into how behavior changes with varying levels of depression since every subject follows a different model. We introduce a group-personalized model as an alternative for personalization where subgroups of individuals that exhibit similar relationships between their behavioral data and mental states are automatically identified. This leads to plausible groupings since most healthy control individuals and individuals with bipolar disorder and borderline personality disorder have been assigned to different groups, reflecting the relationship between different disease categories and different behavioral patterns. However, further validation with more data is needed to assess whether these are “optimal” groupings.

While groupings provide insights, the key challenge for applying group-personalized models in practice remains the determination of which group to allocate new individuals to. In the results above, up to 8 data points (pairs of behavioral features and actual QIDS scores) were used to allocate each individual to a group based on the performance of QIDS score estimation using the model for each group. This provided a significant improvement over the population-level model and a smaller, but still mildly significant, improvement over the fully personalized model trained on the same data used for calibration. In practice, the main advantage of grouping might be to enable better understanding of patient characteristics. While the fully personalized model may take any form, the group-personalized model is restricted to a known set of behavioral phenotypes. With enough individuals available to train models, an exhaustive set of behavioral phenotypes can be obtained. Matching individuals to groups based purely on the predictive performance of the extracted groups on calibration data, as done here, is a naïve approach, and using other indicators of similarity, such as demographics or similarity of sensor data, may improve group-matching when only limited calibration data are available.

The fact that a high number of groups relative to the number of individuals were found suggests that there is indeed a large amount of interindividual variability between subjects. Having access to larger datasets might also increase the number of subjects belonging to each group. With the current dataset, groups with clearly different behavioral patterns could be identified, as shown in Figure 6. Significant improvements in performance of estimating levels of depression from objective, geolocation-derived features were shown, demonstrating further the appropriateness of the groupings, but also the utility of using geolocation as an objective marker for mental health.

Quick Inventory of Depressive Symptomatology Estimation Model

The presented model uses standard linear regression as the core model within groups. This assumes a linear relationship between the model’s predictor(s) and output. Other authors, for instance, Abdullah et al [30], have used more advanced methods such as support vector regression in a similar application, which enables nonlinearity in the relationship between the behavioral features and output variable of interest to be modeled. In the present model, nonlinearity could be modeled using a generalized linear model in place of the standard linear regression. However, the individuals highlighted in the features shown in Figure 4 do not indicate that nonlinearity is a major limitation, but rather variability in the output. For this reason, incorporating nonlinearity may exacerbate any overfitting of the model to the training data available.

Limitations

While the results presented here demonstrate the utility of group-personalized models of behavior to improve regression performance—which might be a very useful approach beyond this application—a number of important limitations need to be discussed.

First, a limitation of this work, in common with many previous studies of objective markers of mental health, is reliance on and comparison with subjective proxies of mental state. In this work, the patient-reported QIDS questionnaire was used to train and evaluate models. Similar patient-reported questionnaires have been used in most previous work. Some studies use clinician-reported measures, but these are still fundamentally subjective proxies and are usually not available at a high sample rate. While this is an accepted limitation of the current work, as more longitudinal data become available for analysis, the properties of the behavioral phenotypes found may in turn help inform our understanding of mental illness.

The model presented here also assumes that individuals always remain in the same group. In reality, individuals may exhibit temporal variability in their response to illness. For example, in some individuals, improvements in mental state may follow different models to deterioration, or the same individual may have different relationships with behavioral features during different episodes.

The main area of interest in using objective markers to monitor levels of mental illness is the transition between states. Detailed investigation of state transitions compared with stable states may provide useful data about which variations in behavior are normal for an individual in remission (as in bipolar disorder) and which variations significantly correlate with the onset of illness episodes. Separating the two will always be crucial for predictive models to perform adequately. A limitation of the data used to train the models presented in this work is that most individuals did not exhibit both stability and variability in illness. Indeed, most individuals were relatively stable in their levels of depression. Again, as more data become available, more complex analysis can be performed.

Conclusions

This paper has demonstrated the limitations of using population-level models to estimate levels of mental illness from behavioral features. Population-level models do not account for natural interindividual variability in how individuals' behavior changes in response to mental illness such as depression. On the other end of the spectrum, fully personalized models built using training data only from specific individuals limits interpretation into clinical phenotypes.

Group-personalized models were therefore presented as a way to augment limited training data available for an individual with data from a group of other individuals who have a similar relationship between their behavior and mental state. Predicting levels of self-reported depression from geolocation-derived features demonstrated the model's appropriateness. Several previous studies have shown the need for personalized modeling for mental health applications due to the high noise levels in behavioral data, as also demonstrated in this work. While there is a clear advantage in using group-personalized models over population-level models, further work must validate these models. Optimal group allocation remains an open question, but the value in the interpretability of the grouped models has been demonstrated. As further data are collected, the utility of the model is expected to increase because more refined models can be inferred from the groups of individuals found.

Acknowledgments

This work was supported by Engineering and Physical Sciences Research Council and Wellcome Trust Centre Grants 098461/Z/12/Z (Sleep, Circadian Rhythms & Neuroscience Institute) and 102616/Z (Collaborative Network for Bipolar Research to Improve Outcomes). NP and OC acknowledge the support of the Research Councils UK Digital Economy Programme grant number EP/G036861/1 (Oxford Centre for Doctoral Training in Healthcare Innovation). KEAS was supported by the National Institute for Health Research Oxford Health Biomedical Research Centre. All data underpinning this paper are available via the University of Oxford data repository [34].

All authors reviewed and approved the manuscript for publication. The views expressed in this publication are those of the authors and not necessarily those of the National Institute for Health Research, the National Health Service, or the Department of Health.

Conflicts of Interest

GMG is a National Institute for Health Research Senior Investigator; received honoraria for speaking or chairing educational meetings from Angelini, Lundbeck (/Otsuka or /Takeda), Medscape, Pfizer, Servier, Shire, Sun Pharma; advised Allergan, Compass pathways, MSD, Lundbeck (/Otsuka or /Takeda), Minervra, Servier, Shire; and holds shares in P1vital.

References

1. Wielopolski J, Reich K, Clepce M, Fischer M, Sperling W, Kornhuber J, et al. Physical activity and energy expenditure during depressive episodes of major depression. J Affect Disord 2015 Mar 15;174:310-316. [doi: 10.1016/j.jad.2014.11.060] [Medline: 25532078]
2. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S, et al. StudentLife: Assessing mental health, academic performance and behavioral trends of college students using smartphones. New York: ACM; 2014 Presented at: UbiComp '14: ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 13-17, 2014; Seattle, USA p. 3-14. [doi: 10.1145/2632048.2632054]
3. Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, et al. Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. J Med Internet Res 2015 Jul 15;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26180009]
4. Mehrotra A, Hendley R, Musolesi M. Towards Multi-modal Anticipatory Monitoring of Depressive States through the Analysis of Human-Smartphone Interaction. New York: ACM; 2016 Presented at: UbiComp '16: ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct; September 12-16, 2016; Heidelberg, Germany p. 1132-1138. [doi: 10.1145/2968219.2968299]
5. Gruenerbl A, Osmani V, Bahle G, Carrasco JC, Oehler S, Mayora O, et al. Using smart phone mobility traces for the diagnosis of depressive and manic episodes in bipolar patients. New York: ACM; 2014 Presented at: AH '14: 5th Augmented Human International Conference; March 07-08, 2014; Kobe, Japan. [doi: 10.1145/2582051.2582089]

6. Grünerbl A, Muaremi A, Osmani V, Bahle G, Ohler S, Tröster G, et al. Smartphone-based recognition of states and state changes in bipolar disorder patients. IEEE J Biomed Health Inform 2015 Jan;19(1):140-148. [doi: 10.1109/JBHI.2014.2343154] [Medline: 25073181]
7. Saeb S, Lattie EG, Schueller SM, Kording KP, Mohr DC. The relationship between mobile phone location sensor data and depressive symptom severity. PeerJ 2016;4:e2537 [FREE Full text] [doi: 10.7717/peerj.2537] [Medline: 28344895]
8. Mohr DC, Zhang M, Schueller SM. Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning. Annu Rev Clin Psychol 2017 May;13:23-47. [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]
9. Berrouiguet S, Perez-Rodriguez MM, Larsen M, Baca-García E, Courtet P, Oquendo M. From eHealth to iHealth: Transition to Participatory and Personalized Medicine in Mental Health. J Med Internet Res 2018 Jan 03;20(1):e2 [FREE Full text] [doi: 10.2196/jmir.7412] [Medline: 29298748]
10. Canzian L, Musolesi M. Trajectories of Depression: Unobtrusive Monitoring of Depressive States by means of Smartphone Mobility Traces Analysis. New York: ACM; 2015 Presented at: UbiComp '15: ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 07-11, 2015; Osaka, Japan p. 1293-1304. [doi: 10.1145/2750858.2805845]
11. Lara Ó, Labrador MA. A Survey on Human Activity Recognition using Wearable Sensors. IEEE Commun Surv Tutor 2013;15(3):1192-1209. [doi: 10.1109/SURV.2012.110112.00192]
12. Servia-Rodríguez S, Rachuri KK, Mascolo C, Rentfrow PJ, Lathia N, Sandstrom GM. Mobile Sensing at the Service of Mental Well-being: a Large-scale Longitudinal Study. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee; 2017 Presented at: WWW '17: 26th International Conference on World Wide Web; April 03-07, 2017; Perth, Australia p. 103-112. [doi: 10.1145/3038912.3052618]
13. Hong J, Ramos J, Dey AK. Toward Personalized Activity Recognition Systems With a Semipopulation Approach. IEEE Trans Hum Mach Syst 2016 Feb;46(1):101-112. [doi: 10.1109/THMS.2015.2489688]
14. Matthews M, Murnane E, Snyder J, Guha S, Chang P, Doherty G, et al. The double-edged sword: A mixed methods study of the interplay between bipolar disorder and technology use. Comput Human Behav 2017 Oct;75:288-300. [doi: 10.1016/j.chb.2017.05.009]
15. Lane ND, Xu Y, Lu H, Hu S, Choudhury T, Campbell AT, et al. Enabling Large-scale Human Activity Inference on Smartphones using Community Similarity Networks (CSN). 2011 Presented at: UbiComp '11: 13th International Conference on Ubiquitous Computing; September 17-21, 2011; Beijing, China p. 355-364. [doi: 10.1145/2030112.2030160]
16. Abdullah S, Lane ND, Choudhury T. Towards Population Scale Activity Recognition: A Framework for Handling Data Diversity. Palo Alto: AAAI Press; 2012 Jul 14 Presented at: Twenty-Sixth AAAI Conference on Artificial Intelligence; July 22-26, 2012; Toronto, Canada URL: https://www.aaai.org/ocs/index.php/AAAI/AAAI12/paper/view/5169
17. Palmius N, Osipov M, Bilderbeck AC, Goodwin GM, Saunders K, Tsanas A, et al. A multi-sensor monitoring system for objective mental health management in resource constrained environments. : IET; 2014 Presented at: AHT 2014: Appropriate Healthcare Technologies for Low Resource Settings; September 17-18, 2014; London, UK. [doi: 10.1049/cp.2014.0764]
18. Palmius N, Tsanas A, Saunders KEA, Bilderbeck AC, Geddes JR, Goodwin GM, et al. Detecting Bipolar Depression From Geographic Location Data. IEEE Trans Biomed Eng 2017 Aug;64(8):1761-1771 [FREE Full text] [doi: 10.1109/TBME.2016.2611862] [Medline: 28113247]
19. Carr O, Saunders KEA, Tsanas A, Bilderbeck AC, Palmius N, Geddes JR, et al. Variability in phase and amplitude of diurnal rhythms is related to variation of mood in bipolar and borderline personality disorder. Sci Rep 2018 Jan 26;8(1):1649 [FREE Full text] [doi: 10.1038/s41598-018-19888-9] [Medline: 29374207]
20. Carr O, Saunders KEA, Bilderbeck AC, Tsanas A, Palmius N, Geddes JR, et al. Desynchronization of diurnal rhythms in bipolar disorder and borderline personality disorder. Transl Psychiatry 2018 Apr 12;8:79 [FREE Full text] [doi: 10.1038/s41398-018-0125-7] [Medline: 29643339]
21. Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, et al. The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), Clinician Rating (QIDS-C), and Self-Report (QIDS-SR): A Psychometric Evaluation in Patients with Chronic Major Depression. Biol Psychiatry 2003 Sep 01;54(5):573-583. [doi: 10.1016/S0006-3223(02)01866-8] [Medline: 12946886]
22. Tsanas A, Saunders KEA, Bilderbeck AC, Palmius N, Osipov M, Clifford GD, et al. Daily longitudinal self-monitoring of mood variability in bipolar disorder and borderline personality disorder. J Affect Disord 2016 Nov 15;205:225-233 [FREE Full text] [doi: 10.1016/j.jad.2016.06.065] [Medline: 27449555]
23. Tsanas A, Saunders K, Bilderbeck A, Palmius N, Goodwin G, De Vos M. Clinical Insight Into Latent Variables of Psychiatric Questionnaires for Mood Symptom Self-Assessment. JMIR Ment Health 2017 May 25;4(2):e15 [FREE Full text] [doi: 10.2196/mental.6917] [Medline: 28546141]
24. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics). New York: Springer; 2009.
25. Tibshirani R. Regression Shrinkage and Selection via the Lasso. J R Stat Soc Series B Methodol 1996;58(1):267-288.
26. Osborne MR, Presnell B, Turlach BA. On the LASSO and its Dual. J Comput Graph Stat 2000;9(2):319-337. [doi: 10.1080/10618600.2000.10474883]
27. Müller P, Quintana FA, Jara A, Hanson T. Bayesian Nonparametric Data Analysis (Springer Series in Statistics). Cham: Springer International Publishing; 2015.

28. Park T, Casella G. The Bayesian Lasso. J Am Stat Assoc 2008;103(482):681-686. [doi: 10.1198/016214508000000337]
29. Neal RM. Markov Chain Sampling Methods for Dirichlet Process Mixture Models. J Comput Graph Stat 2000;9(2):249-265. [doi: 10.1080/10618600.2000.10474879]
30. Abdullah S, Matthews M, Frank E, Doherty G, Gay G, Choudhury T. Automatic detection of social rhythms in bipolar disorder. J Am Med Inform Assoc 2016 May 01;23(3):538-543 [FREE Full text] [doi: 10.1093/jamia/ocv200] [Medline: 26977102]
31. Andoni A, Indyk P. Near-Optimal Hashing Algorithms for Approximate Nearest Neighbor in High Dimensions. Commun ACM 2008 Jan;51(1):117-122. [doi: 10.1145/1327452.1327494]
32. Drucker H. Improving Regressors Using Boosting Techniques. 1997 Presented at: ICML '97: Fourteenth International Conference on Machine Learning; July 8-12, 1997; Nashville, USA p. 107-115.
33. Wallach HM, Jensen ST, Dicker L, Heller KA. An Alternative Prior Process for Nonparametric Bayesian Clustering. 2010 Presented at: AISTATS 2010: Thirteenth International Conference on Artificial Intelligence and Statistics; May 13-15, 2010; Chia, Italy p. 892-899 URL: https://arxiv.org/abs/0801.0461
34. Palmius N, Saunders K, Carr O, Geddes J, Goodwin G, De Vos M. Data underpinning paper group-personalized regression models for prediction of mental health scores from objective mobile phone data streams. Oxford University Research Archive 2018. [doi: 10.5287/bodleian:Bej7r6y0d]

Abbreviations

AMoSS: automated monitoring of symptom severity
BD: bipolar disorder
BDP: borderline personality disorder
DP: Dirichlet process
HC: healthy control
MAE: mean absolute error
QIDS: 16-item Quick Inventory of Depressive Symptomatology, Self-Report version

Edited by J Torous; submitted 22.02.18; peer-reviewed by D Mohr, S Abdullah; comments to author 04.04.18; revised version received 29.05.18; accepted 10.06.18; published 22.10.18.

Please cite as:
Palmius N, Saunders KEA, Carr O, Geddes JR, Goodwin GM, De Vos M
Group-Personalized Regression Models for Predicting Mental Health Scores From Objective Mobile Phone Data Streams: Observational Study
J Med Internet Res 2018;20(10):e10194
URL: https://www.jmir.org/2018/10/e10194/
doi:10.2196/10194
PMID:30348626

Original Paper

Feasibility and Acceptability of Mobile Phone–Based Auto-Personalized Physical Activity Recommendations for Chronic Pain Self-Management: Pilot Study on Adults

Mashfiqui Rabbi[1*], PhD; Min SH Aung[2*], PhD; Geri Gay[2], PhD; M Cary Reid[3], MD, PhD; Tanzeem Choudhury[2], PhD

[1]Department of Statistics, Harvard University, Cambridge, MA, United States
[2]Information Science Department, Cornell University, Ithaca, NY, United States
[3]Department of Medicine, Weill Cornell Medical College, New York, NY, United States
[*]these authors contributed equally

Corresponding Author:
Mashfiqui Rabbi, PhD
Department of Statistics
Harvard University
1 Oxford Street
Cambridge, MA,
United States
Phone: 1 603 667 1797
Email: mrabbi@fas.harvard.edu

Abstract

Background: Chronic pain is a globally prevalent condition. It is closely linked with psychological well-being, and it is often concomitant with anxiety, negative affect, and in some cases even depressive disorders. In the case of musculoskeletal chronic pain, frequent physical activity is beneficial. However, reluctance to engage in physical activity is common due to negative psychological associations (eg, fear) between movement and pain. It is known that encouragement, self-efficacy, and positive beliefs are effective to bolster physical activity. However, given that the majority of time is spent away from personnel who can give such encouragement, there is a great need for an automated ubiquitous solution.

Objective: MyBehaviorCBP is a mobile phone app that uses machine learning on sensor-based and self-reported physical activity data to find routine behaviors and automatically generate physical activity recommendations that are similar to existing behaviors. Since the recommendations are based on routine behavior, they are likely to be perceived as familiar and therefore likely to be actualized even in the presence of negative beliefs. In this paper, we report the preliminary efficacy of MyBehaviorCBP based on a pilot trial on individuals with chronic back pain.

Methods: A 5-week pilot study was conducted on people with chronic back pain (N=10). After a week long baseline period with no recommendations, participants received generic recommendations from an expert for 2 weeks, which served as the control condition. Then, in the next 2 weeks, MyBehaviorCBP recommendations were issued. An exit survey was conducted to compare acceptance toward the different forms of recommendations and map out future improvement opportunities.

Results: In all, 90% (9/10) of participants felt positive about trying the MyBehaviorCBP recommendations, and no participant found the recommendations unhelpful. Several significant differences were observed in other outcome measures. Participants found MyBehaviorCBP recommendations easier to adopt compared to the control (β_{int}=0.42, $P<.001$) on a 5-point Likert scale. The MyBehaviorCBP recommendations were actualized more (β_{int}=0.46, $P<.001$) with an increase in approximately 5 minutes of further walking per day (β_{int}=4.9 minutes, $P=.02$) compared to the control. For future improvement opportunities, participants wanted push notifications and adaptation for weather, pain level, or weekend/weekday.

Conclusions: In the pilot study, MyBehaviorCBP's automated approach was found to have positive effects. Specifically, the recommendations were actualized more, and perceived to be easier to follow. To the best of our knowledge, this is the first time an automated approach has achieved preliminary success to promote physical activity in a chronic pain context. Further studies are needed to examine MyBehaviorCBP's efficacy on a larger cohort and over a longer period of time.

(J Med Internet Res 2018;20(10):e10147) doi:10.2196/10147

KEYWORDS

chronic pain; machine learning; personalization; chronic back pain; reinforcement learning

Introduction

Background

Chronic pain is defined as pain that persists despite the resolution of injury or pathology [1] and attributed to changes in the central and peripheral nervous system resulting in amplified or uninhibited pain signals [2,3]. Persistent pain over long periods of time can affect people both physically and psychologically. Chronic pain is also closely linked with distress, affect behavior, and loss of productivity. Some studies have shown that 86% of people with chronic pain report difficulties with sleep [4], 70% have trouble concentrating [4], 44% to 51% report anxiety [5], and 88% express anger due to not seeing improvements [5]. Moreover, productivity losses can also be significant, with 4.6 hours per week rising to 5.5 hours per week lost due to chronic back pain [6]. These adverse effects can also result in other comorbidities such as depression [5,7-9]. In terms of global prevalence, the World Health Organization recognizes chronic pain as a public health problem worldwide [10] with a prevalence of 32% in low-income countries and 30% in high-income countries [11,12]. In the United States alone, 120 million adults are reported to be suffering from chronic pain [13] with related costs exceeding even that of diabetes, cancer, and heart disease [14]. Chronic pain is also related to substance abuse, with 12 million people (aged 12 years and older) reporting nonmedical use of pain medication, and overdose-related deaths are rising annually [15].

A particularly common form of chronic pain is of a musculoskeletal nature, which affects 1 in 10 adults globally. This form of chronic pain is also a leading cause of disability, with 28% reporting limitations in movement due to the condition [16]. That said, it is well understood that successful management of musculoskeletal chronic pain is achievable with regular and sustained physical activity [10]. This is due to activity having the effect of protecting against muscle weakening and inhibiting the neurophysiological mechanisms underlying the spread of pain [4]; in addition, physical activity does not have the issue of side effects that come with the consumption of pharmaceuticals [10,17]. Indeed, several recent reviews and guidelines from the US Centers for Disease Control and Prevention and UK National Institute for Health and Care Excellence have strongly encouraged clinicians to prescribe nonpharmacologic approaches that include movement-based therapies [10,18-20].

Despite the benefits of physical activity, adherence to regular and sustained physical activity is low [12,21]. This is principally due to perceived pain exacerbation from activity, which over time results in negative psychological associations (eg, fear) between movement and pain. Some of the psychological associations, typically fear and anxiety [22,23], manifest at an affective level [5] and are closely linked to more cognitive associations such as negative beliefs [5] and catastrophization [24], where subjective interpretation of pain severity may make the pain seem worse than it is [25,26]. These negative associations can lead to a reluctance to continue and even avoidance of activity and therapies [23,27] that in turn can lead to lowered self-efficacy to engage in preventative measures [24]. Lack of engagement over time can further result in weakening, disability, or even impairment in motor control where there is fproprioceptive dysfunction [28].

Low adherence to physical activity is further compounded by the need to self-manage. Typically, in day-to-day life settings, there is no care provider present to offer encouragement and guidance [12,29]. A variety of mobile phone apps have been proposed in recent years for chronic pain self-management [12,30-32]; however, the majority of mobile health (mHealth) apps to date for chronic pain do not leverage the available array of onboard sensors [33,34]. In this paper, we argue that sensor data can offer a new method that makes self-management of physical activity easier for chronic pain. Self-management becomes easier because suggestions are generated and adapted automatically from sensor data, so participants do not need to manually manage or track the suggestions they want to follow. We further reduce self-management burden by prioritizing features that make the suggestions more actionable within the psychological barriers of chronic pain [35,36]. Thus, participants do not need to use trial and error to figure out which suggestions are more or less actionable for chronic pain.

Strategies to Address Psychological Barriers of Chronic Pain

While the utility of physical exercise for chronic pain is well known, it has also been found that introduction of new exercise tasks is more successful when small changes are made to current daily activities [12]. Other studies advocate a shift of attention strategy [37] as well as focusing on pleasurable activities [38]. These studies suggest the need to base any recommendation system on personal preferences and routine behavior. Given these lessons, we aimed to use mobile phone sensors as a natural way to acquire data on habitual behavior with respect to activity preferences. We also aimed to use this data as a basis from which new activity suggestions can be issued that are based on the contextual information taken from the tracked data. Such suggestions will likely be perceived as familiar. Also, given that the users know they have done similar activities before, this may lead to more positive beliefs and a better sense of self-efficacy. Thus, the suggestions themselves could be perceived as less effortful compared to fixed text message–based approaches, where messages are constructed before a study and the message contents do not adapt to routines or preferences of study participants [39].

A Personalized, Self-Efficacious, and Low-Effort Suggestion Engine

We developed MyBehaviorCBP, a mobile phone app that operationalizes various strategies to address psychological barriers of chronic pain. MyBehaviorCBP uses machine learning on sensor data and self-reported physical activity logs and automatically generates physical activity recommendations based on an individual's past behavior. This strategy of persuasion has been shown to be effective in MyBehavior, our predicating system designed for general populations [35,36]. This earlier app was designed to promote more energetic exercising and lower calorie dietary intake based on the user's past actions. MyBehavior was shown to have affected a sustained positive behavior change within a general population in a 14-week pilot study. In this study, we repurposed this system and developed MyBehaviorCBP specifically for individuals with chronic pain. In MyBehaviorCBP, the user's mobility state (episodes of walking or stationary state inferred from the accelerometer signals) and geolocation are passively tracked without the need for any user input. Activities that cannot be captured with sensors are entered manually. The next stage is to find recurring patterns within the physical activity data of each participant. Once recurring patterns are established, a set of new recommendations is issued based on these recurring patterns with small changes applied. We focus on these recurring patterns to generate new suggestions because participants have likely done similar actions before. Furthermore, the app uses the tracked movement data as a way to monitor which recommendations were actually preferred and executed. Subsequent recommendations are adapted based on the most acted upon previous recommendations. To the best of our knowledge, MyBehaviorCBP is the first app to promote physical activity with an automated data-driven approach in the chronic pain context.

Objective of This Study

In this paper, we report on a formative study on the use of MyBehaviorCBP and present results from a 5-week pilot study among individuals with chronic back pain (N=10). Since the MyBehaviorCBP automated suggestion generation approach is being tested for the first time in the chronic pain context, we investigate the feasibility and acceptability of the approach before an expensive randomized controlled trial. Prior works have recommended small pilot trials (N>4) for novel mHealth apps to investigate early evidence of acceptance and use demonstrating the intervention is affecting the intended outcomes and document lessons learned, if any, for future improvements [40,41]. To this end, we (1) determined whether the MyBehaviorCBP recommendations were perceived as easy and actionable compared to randomly generated recommendations, (2) examined preliminary evidence to see whether the intentions led to an actual increase in physical activity behavior, and (3) solicited participant feedback on using the app to fine-tune future versions of the app.

Methods

Study Design

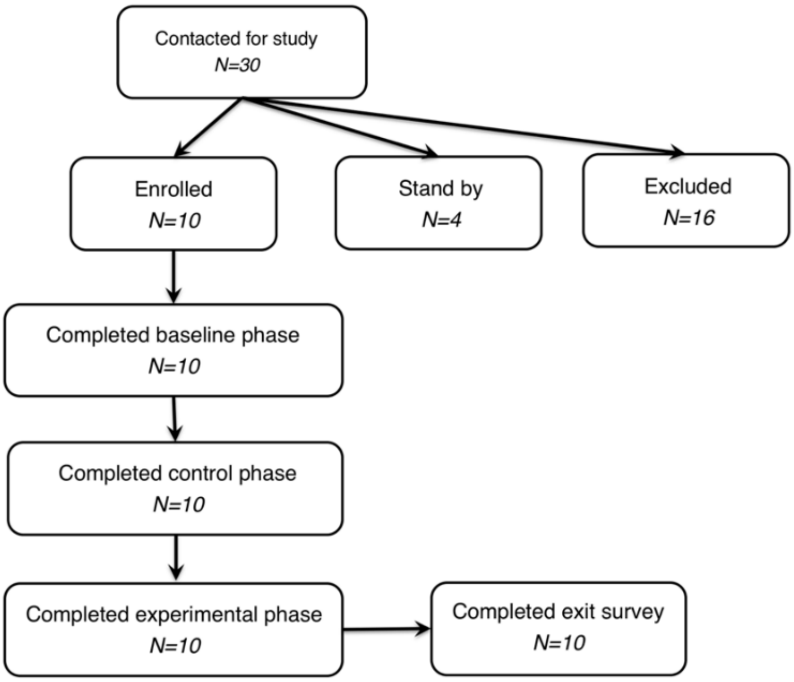
We conducted a 5-week within-subject study on 10 individuals with chronic back pain. The first week of the study served as a baseline period where participants familiarized with the app. No physical activity recommendation was given in the first week. The next 2 weeks were a control phase where 7 suggestions were randomly chosen every day from a pool of suggestions. This pool of suggestions was created by a fitness expert according to the US National Institutes of Health guidelines for healthy living [42,43] and these suggestions were pilot-tested in past studies [35,36]. These suggestions, however, were generic and unrelated to participants' past behaviors. In weeks 4 and 5, the experimental phase was conducted where participants received MyBehaviorCBP-generated recommendations based on their own behaviors. The study was single-blinded in that only the experimenters were aware of when the different types of suggestions during the control and experimental phases were activated.

During each day of the control and experimental phases, participants filled out a short in-phone survey in the evening. The survey asked about the ease of following recommendations, how many recommendations they followed, and their emotional state. In addition to the daily surveys, participants completed a Web-based exit survey after the study. The exit survey asked about the helpfulness of the recommendations, what future changes they would want to see, and whether they would recommend this app to other people with chronic back pain.

Participants

Given the prevalence of chronic pain, invitation of the study was sent via the Wellness Center and retiree mailing lists from Cornell University. Recruitment was restricted to participants with a history of chronic back pain (≥6 months in duration) and willingness to use MyBehaviorCBP on an Android mobile phone, either their own or one provided by the study. Further inclusion criteria were having some reasonable level of outdoor movement (eg, traveling to and from work), not being significantly housebound, having a basic level of mobile phone proficiency, being between ages 18 and 65 years, and being fluent in English. Exclusion criteria, determined during an initial interview, were the need of mobility aids; having had joint replacement, arthrodesis, or limb amputation; having a learning disability; or being pregnant, but no subject fell into these categories. Eligible participants were invited for a face-to-face session where informed consent was acquired and instructions for using the app were provided (Figure 1 shows the participant flow diagram); 10 participants (3 male, 7 female, aged 31 to 60 years) were recruited, and all participants completed the study. Participants reported a range of causes for their chronic back pain, including herniated disc, rotated vertebrae, scoliosis, sciatica, and tendinitis [44] and had histories of 5 to 33 years of chronic back pain. It should be noted here that independent clinical diagnoses past that of the subjects' own declarations about their respective causes of chronic pain were not applied. Since we are only interested in the chronic pain aspect of their condition rather the specific causes, we followed the approach of self-reporting of pain. As pain is subjective, self-report of the presence, level, and persistence of pain remains the standard assessment approach. After initial contact, some participants were further excluded as our app was limited to the Android operating system and there was a limited number of available replacement phones. Also, some potential participants were excluded because they would not be in their normal daily routines (eg, going on vacation) during the time of the user study. From the initial 30 contacted, 16 were excluded and 10 enrolled with 4 on standby.

Figure 1. Participant flow diagram for MyBehaviorCBP pilot study.



The MyBehaviorCBP Intervention

The MyBehaviorCBP app comprised 2 modules: routine behavior recognition module and recommendation generation module.

Routine Behavior Recognition Module

The first stage of MyBehaviorCBP is to log the physical activities of an individual with a combination of movement sensors (geolocation and accelerometer) along with manual input. Similar recurring activities are then grouped together to find routine behaviors. Specifically, activity states such as walking, running, stationary, and in-vehicle are automatically tracked using movement sensors within the phone; these activities are also tagged with the geographical location [45]. Physical activities that cannot be automatically inferred by phone sensors can be manually recorded using a drop-down menu that contains a searchable preloaded set [46]. Once an activity is logged, it is grouped together with other similar activities that have previously occurred. The method for this grouping is a data-clustering algorithm, details of which have been previously published [35,36]. The main intuition is that the same type of activities will co-occur at similar locations and are therefore assigned to the same cluster. For example, episodes of stationary state that occur in an office will have a similar GPS location and would be grouped together. Stationary episodes at a different location would be in a separate cluster from the “stationary in the office” cluster. Episodes of other mobility states are also grouped separately. For example, episodes of walking from an office to a coffee shop would show a similar trajectory of GPS locations and will be grouped together; in principle this would be a different cluster to walking from the office to a bus stand. For the manually searched and logged activities, similar exercises within the preloaded list are grouped together. Figure 2 shows a few examples of physical activity patterns extracted from 2 users.

Figure 2. Visualization of a user’s movements over a week: (a) heatmap showing the locations where the user is stationary everyday, (b) location traces of frequent walks by the user, and (c) location traces of frequent walks by another user.

In the interest of consistency, we will refer to each of these multifaceted clusters as a “behavior” in the remainder of this paper. Note the clustering process is determined for each participant separately without using data from other participants. Furthermore, the clustering is carried out in the phone, and no location data is exported to the cloud, minimizing privacy risks.

Suggestions Generation Module

Once the tracked data are grouped into different behaviors, the app then uses a sequential decision-making algorithm (multi-armed bandit or MAB [47,48]) to select and rank recommendations that are maximized to be both actionable and beneficial. In the following, we describe the main factors within the data that we consider to be important in the context of chronic pain. Subsequently, we will describe how the algorithm operationalizes these factors as part of its optimization process.

- Most frequent and repeated behaviors are prioritized. In doing this, we aim to exploit the fact that participants are familiar with these frequent behaviors and they likely have a higher level of mastery or sense of self-efficacy toward undertaking those actions [49].
- Less intensive and energetic actions are prioritized. For example, walking is prioritized over running or gym exercises. This factor is considered to promote easier or perceived as easier suggestions, which may be more compelling in situations when there is fear or anxiety of contemplating exercise [39].
- Newly generated suggestions are based on the continuation of small changes made to a user’s existing repeated behaviors. As suggested in Singh et al [12], it is attractive for those with chronic pain if less change to current behavior is needed when adopting a new therapy compared to suggestions that significantly differ from their existing routine.
- Suggestions will be uniquely contextualized to each user. Contextual information such as road or place names (Figure 2) and durational information can be added to the suggestion to further elicit a sense of familiarity [12].

In addition to the main tenets listed above, a further requirement is the need for the system to be adaptive and future proof. Since the suggestions are generated when the app is being used and data acquired, the system only has an account of the user’s past behaviors and the suggestions that have been actualized. This information is incomplete to inform what may happen in the future (eg, an ineffective suggestion from the past may become effective at a later point in time and vice versa). Thus, the system needs to have the capacity to adjust over time and adapt if necessary. Within the MAB framework, principles from the reinforcement learning (RL) branch of artificial intelligence are used, and this learning paradigm is designed to address the task of being continually adaptive. In this context, the RL agent can take a sequence of decisions in an environment to reach a predefined objective where each subsequent decision is based on the success or failure of the previous decisions.

One can consider the MyBehaviorCBP system as an RL agent as follows: let $X=\{x_1,x_2,x_3,\ldots,x_n\}$ denote the set of recommendations where each x_i is related to a user behavior (ie, walking to bus stop, sitting in the office). At the start of day d, the system chooses a subset of X and issues them to its user. At the end of the day, a reward score r_{id} for recommendation x_i is calculated according to the following quantity: $r_{id}=m_{id}\times easiness\ (x_i)$, where m_{id} represents number of minutes a participant spent actually doing x_i during the day d; m_{id} is equal to zero if the recommendation is not adopted. The *easiness* (x_i) function depends on how light or less vigorous the recommendation x_i is (eg, walking is lighter than running or gym exercise). We quantify the lightness of an activity as inverse of metabolic equivalent of task (MET) because vigorous activities tend to have higher MET values [46]. To summarize, the longer a recommendation is followed and the lower MET value it has, the higher reward it receives (eg, a longer walk would receive a higher reward than a shorter walk). But, between behaviors of the same duration but different action, eg, 30 minutes of walking and 30 minutes of gym exercise, walking would receive a higher reward because it is easier with a lower MET value.

One exception to the above equation for r_{id} is when there is a stationary behavior (eg, sitting). This is because being stationary has a low MET value, and a lot of time is generally spent in a sedentary manner. Thus, our prior formulation of r_{id} would give high reward to stationary behavior because both m_{id} and *easiness* are higher. However, rewarding more stationary behavior makes little sense, and MyBehaviorCBP’s goal is to reward more movement. Therefore, MyBehaviorCBP make the following adjustment to the reward of stationary behaviors; for each stationary behavior x_j X, the reward is: $r_{jd}=m_{jd}\times easiness\ (walking)\times 3/60$, where easiness of walking is the inverse of the MET value of walking activity. Defined this way, MyBehaviorCBP would suggest a small change of 3-minute walking breaks for every hour of stationary behavior.

At the end of each day d, 2 quantities for each x_i are computed: mean daily reward, $r^{\wedge}_{id}$ (= $[r_{i1}+r_{i2}+r_{i3}+\ldots+r_{id}]/d$) and mean daily minutes spent $m^{\wedge}_{id}$ (= $[m_{i1}+m_{i2}+m_{i3}+\ldots+m_{id}]/d$). A suggestion with high $r^{\wedge}_{id}$ means the suggestion is easy and repeated often. On day d +1, a new subset of recommendations is chosen with the following composition: 80% with the highest $r^{\wedge}_{id}$ and 20% randomly chosen from the remaining; moreover, the total $m^{\wedge}_{id}$ for the selected recommendations cannot exceed 60 minutes. In this setup, MyBehaviorCBP ensures 80% of the recommendations are easy and have been frequently followed before. For the remaining 20%, we allow for exploration of other recommendations to see if they get adopted. Finally, the limit of 60 minutes is set to aim for a predefined duration of exercise. This is done since the therapeutic objective here is not necessarily to encourage as much activity as possible but to regularly adhere to some predefined target. To operationalize this, we use a MAB variant called the multi-armed bandit with knapsack [50] using an ε-greedy strategy [48]. This would encourage the issuance of easier-to-do recommendations that simply target reaching 60 minutes rather than having no upper bound.

Figure 3 shows examples of MyBehaviorCBP suggestions for 2 different users. The screenshots are distinct and show

MyBehaviorCBP’s capability to personalize to different users. Short encouraging texts are also added to the recommendations, such as (1) continuing existing walking behaviors, (2) taking short walking breaks as a small change to stationary behaviors, or (3) encouraging to continue other exercises. Along with this text, the app shows how many minutes are achievable (equal to $m^{\wedge}_{id}$) and should be done for each of the suggestions.

Measures

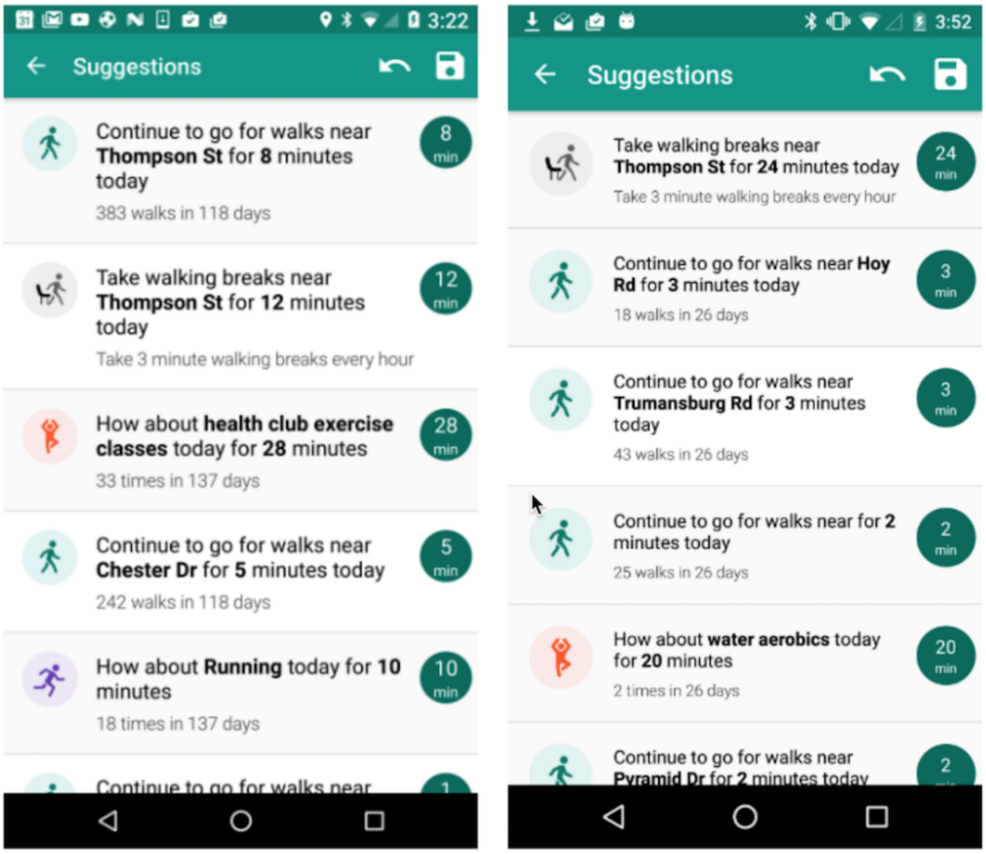
The MyBehaviorCBP system intends to encourage more physical activity over a sustained period of time. However, given the early stage of the technology, a 5-week pilot study was conducted. The goal of the pilot was to investigate the feasibility of MyBehaviorCBP, which was measured by 3 factors: use, acceptability, and early efficacy. In addition, we report lessons learned for future improvements [40,41,51]. These factors are measured with a combination of phone logs, daily evening surveys during the study, and an exit survey (see Table 1 for measured outcomes). Phone logs contain passively collected records of whether MyBehaviorCBP was opened and the type of physical activity (ie, walking, stationary) for every minute during the study. The daily evening surveys were done in the phone each day of the 5-week study; in these daily evening surveys, we asked about relative ease of the recommendations received, how many recommendations they followed, and their emotional state during the day. After the 5-week study, an exit survey was conducted on the Web that asked about the helpfulness of the recommendations, what future changes they would want to see, and whether they would recommend this app to other people with chronic back pain.

Use was measured by how frequently study participants opened the app, recorded from the phone log. Acceptability, a more complex quantity, was measured by traingulating a variety of self-reports that focused on intention and behavior toward the recommendations. We specifically measured perceived easiness, which indicates the actionability of the recommendations [39,49], intention, behavior, and helpfulness. Early efficacy is another complex quantity we measured with several proximal measures that can lead to reduction of pain in the long term [52]. Finally, open-ended questions were used to pin down which features participants found useful and what features they saw as missing; this data will be used to map out future refinements of MyBehaviorCBP.

Analysis Plan

Number of times the app is accessed was analyzed using the simple descriptive statistics of mean and standard deviation. The acceptability and early efficacy outcomes are less straight forward to analyze because data points from the same subject being likely correlated and different subjects having different baseline conditions at the start of the study (eg, different levels of physical activity and type of chronic back pain) [53,54]. Therefore, we apply linear mixed-effect models to adjust for repeated measures and intersubject variability [53].

Figure 3. MyBehaviorCBP’s personalized suggestions for 2 users.



Continue to go for walks near **Thompson St** for **8** minutes today
383 walks in 118 days
8 min

Take walking breaks near **Thompson St** for **12** minutes today
Take 3 minute walking breaks every hour
12 min

How about **health club exercise classes** today for **28** minutes
33 times in 137 days
28 min

Continue to go for walks near **Chester Dr** for **5** minutes today
242 walks in 118 days
5 min

How about **Running** today for **10** minutes
18 times in 137 days
10 min

Continue to go for walks near
1

Take walking breaks near **Thompson St** for **24** minutes today
Take 3 minute walking breaks every hour
24 min

Continue to go for walks near **Hoy Rd** for **3** minutes today
18 walks in 26 days
3 min

Continue to go for walks near **Trumansburg Rd** for **3** minutes today
43 walks in 26 days
3 min

Continue to go for walks near for **2** minutes today
25 walks in 26 days
2 min

How about **water aerobics** today for **20** minutes
2 times in 26 days
20 min

Continue to go for walks near **Pyramid Dr** for **2** minutes today
2

Table 1. Different outcome measures captured in the MyBehaviorCBP pilot study and their purposes.

Data collection methods and description of outcome measure	Purpose of outcome measure
App use	
Record of how many times the app is opened	Use
Physical activity log	
Number of minutes spent walking per day	Early efficacy
Number of minutes spent in nonwalking exercises per day	Early efficacy
Daily evening survey	
Perceived easiness: How easy did today’s suggestions seem after reading them? (Likert scale: 1=I could never do these suggestions to 7=I could always do these suggestions)	Acceptability
Intention: How many suggestions did you want to follow today? (integer value between 0 and 7)	Acceptability
Behavior: How many suggestions did you follow today? (integer value between 0 and 7)	Acceptability
Pain level: Please indicate your pain level today. (Likert scale: 0=no pain to 10=extreme pain)	Early efficacy
Exit survey	
Did receiving suggestions from your phone help you to be more active? (multiple choice: not helpful, somewhat helpful, very helpful)	Acceptability
How likely are you to recommend this app to another person with back pain? (multiple choice: not likely, somewhat likely, very likely)	Acceptability
What changes do you think could be made to the app that would make it more effective in helping you be more active? (open-ended)	Future improvement

The type of intervention is considered as a fixed effect, and we coded the intervention type as 0 and 1 for control (ie, the static suggestions generated by experts) and experimental phases (ie, MyBehaviorCBP suggestions), respectively. Coded this way, the intervention coefficient would represent the relative improvement of the outcome measure of MyBehaviorCBP over the control. When we included time (as day within the study) as a fixed effect, it was found to be not significant. Also, we tested the study participant identity as a random effect and found it to be significant in likelihood ratio tests ($P<.01$) for all outcome measures [53], which means significant interpersonal variability exists and a mixed-effect model is necessary. The parameter estimation for all models was generated using maximum likelihood estimation [53]. In addition, we computed effect size measures by dividing the mean difference of an outcome measure between the control and experiment phases with pooled standard deviation [55]. Finally, the open-ended questions in the exit survey are broken down into themes using thematic analysis [56].

Results

Use

Over the 5-week study with 10 participants, the mean number of times the MyBehaviorCBP app was opened is 106.9 during the control and experiment phases (μ=106.9, σ=56.9, q_{25}=76.1, q_{50}=89.6, q_{75}=105.5), which is 3.2 times on average per day. Figure 4 shows the average number of times a participant opened the MyBehaviorCBP app over time. For both control (ie, static suggestions) and experiment phases (ie, MyBehaviorCBP-generated suggestions), there is an initial period of high use but over time the use decreased.

Acceptability of the Suggestions

In the exit survey, the participants reacted positively about MyBehaviorCBP recommendations, with 2 of 10 participants finding MyBehaviorCBP recommendations very helpful and 8 of 10 finding MyBehaviorCBP recommendations somewhat helpful. No participant reported the recommendations unhelpful. All participants (10/10) reported that they would likely recommend the app to other people with chronic back pain.

The acceptability of MyBehaviorCBP was also measured using (1) self-reported rating of easiness of the recommendations, (2) how many recommendations the participants wanted to follow, and (3) how many recommendations the participants actually followed. The results of the statistical analysis are reported in Table 2, and Figure 5 shows the mean and standard deviations. In the interest of convenience, only important statistics are shown in Table 2. The intervention coefficients (β_{int}) are reported along with P values (P_{int}) and 95% confidence intervals. We observe there were significant changes of β_{int} for number of recommendations followed (β_{int}=0.46, $P<.001$). In real terms, participants within the experimental phase adopted 1 extra recommendation every 2 days. On the other hand, participants wanted to follow the control phase recommendations more than the experimental phase (β_{int}=–0.2, P=.02), which means control group recommendations were effective to increase intention but not for actualizing a recommendation. Regarding easiness, participants perceived the MyBehaviorCBP recommendations to be easier than the control, which means they were perceived as low-burden (β_{int}=0.42, $P<.001$). This is important in the chronic pain population, which is often reluctant to be physically active.

The number of self-reported recommendations followed and wanted to follow, however, had important differences for different emotional states in the day. Figure 6 shows the means for several outcomes under different emotional states captured through the PAM [57]. Generally, participants adopted or wanted to adopt more recommendations during positive emotional states. However, during high positive emotional states, generic recommendations were perceived as easy and participants wanted to follow them more, but the MyBehaviorCBP recommendations were adopted more. During negative emotional states, the number of recommendations participants wanted to follow and actually followed are both higher than the control group recommendations.

Figure 4. Number of times a day MyBehaviorCBP app was accessed.

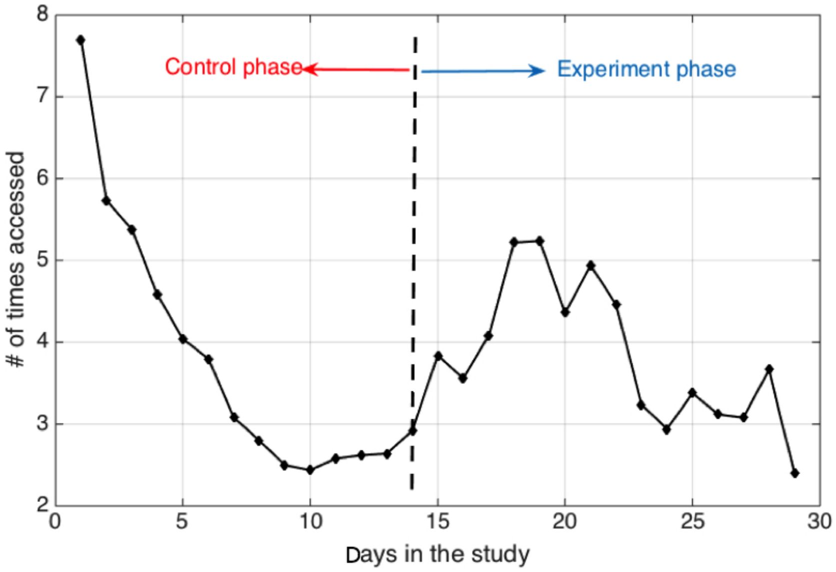


Table 2. Summary of differences between control and MyBehaviorCBP as collected from survey and physical activity logs.

Outcome measure	β_{int}	P_{int}	95% CI_{int}	*d*	–2logL	AIC[a]	BIC[b]	LR[c]
How easy were the suggestions	0.42	<.005	0.2 to 0.6	0.25	817.5	879	894.6	0.009
# of suggestions followed	0.46	<.005	0.2 to 0.7	0.11	4795	4809	4839	0.01
# of suggestions wanted to follow	–0.2	.02	–0.5 to –0.1	–0.2	4795	4809	4839	0.002
Walked (minutes/day)	4.9	.02	0.8 to 8.9	0.31	2123	2131	2144	0.009
Exercised (minutes/day)	9.5	.31	–6.3 to 21.8	0.03	2986	2993	3008	0.01
Pain level	–0.19	.24	–0.5 to 0.14	0.17	1160	1168	1183	0.001

[a] AIC: Akaike information criterion.

[b] BIC: Bayesian information criterion.

[c] LR: likelihood ratio test between the fitted models compared to unconditional mean models [35,53].

Figure 5. Mean and standard deviations of acceptability measures.

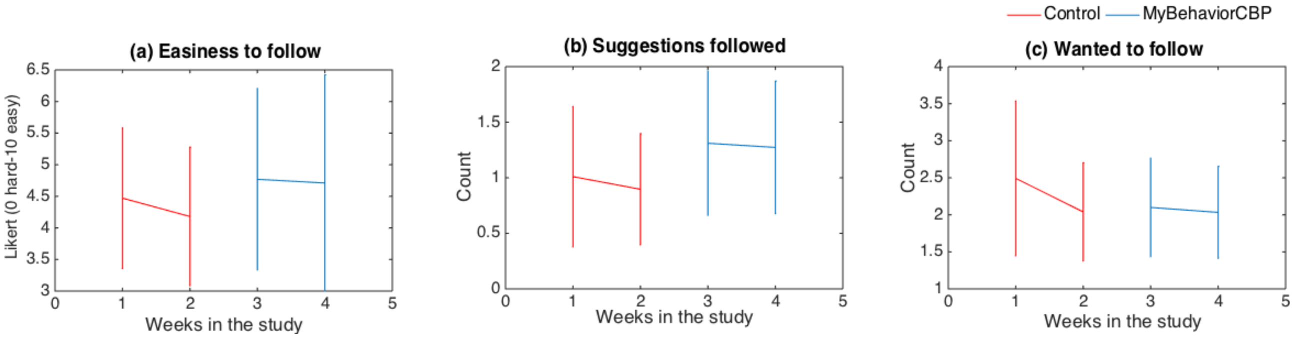


Figure 6. Means of several outcome measures for different emotional states.

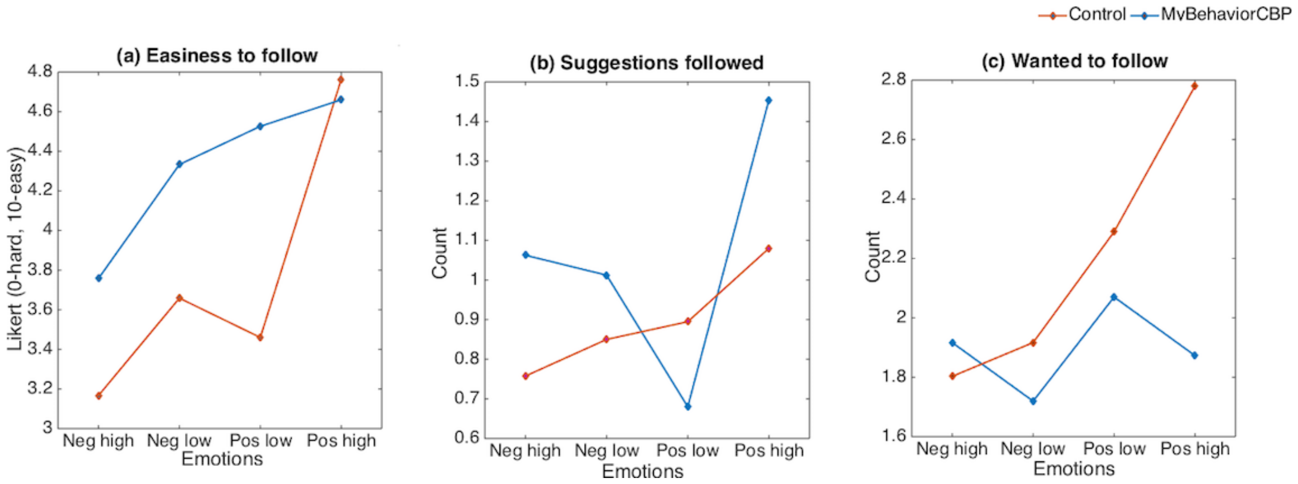
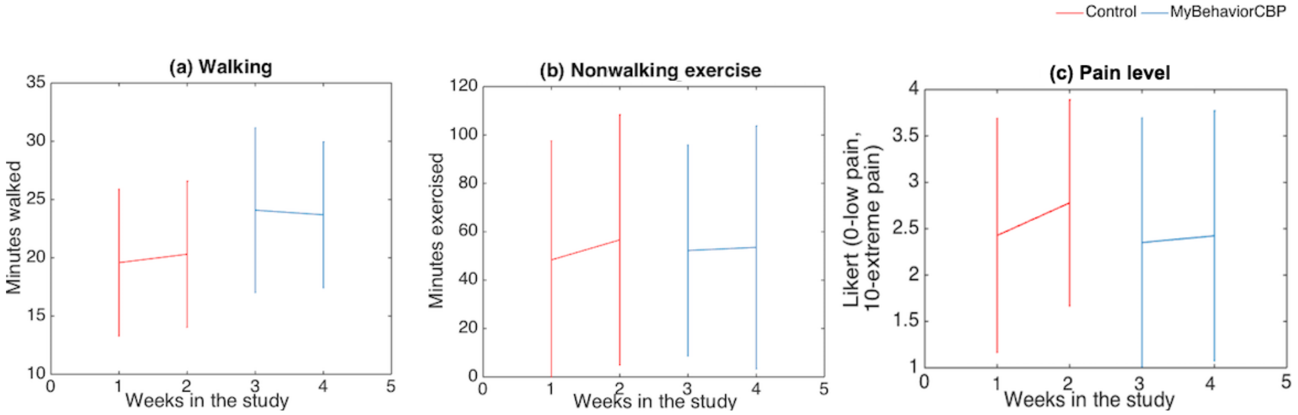


Figure 7. Mean several outcome measures for different preliminary efficacy outcomes.



Preliminary Efficacy on Increasing Physical Activity and Reducing Pain

From Table 2 and Figure 7, the intervention coefficient for daily walked minutes (β_{int}=4.9, *P*=.02) indicates that participants walked for a further 4.9 minutes per day compared to the control phase. β_{int} for nonwalking exercise (β_{int}=9.5, *P*=.31) indicated that participants in the experimental phase exercised for a further 9.5 minutes, but the difference was not significant. Regarding pain level, there was a small nonsignificant 0.19 unit of reduction of pain (β_{int}=–0.19, *P*=.24).

Qualitative Feedback and Lessons Learned

Participants provided qualitative feedback in the exit survey, which provided further insights about the quantitative results and gave directions for future changes. For instance, when we asked participants to compare control group recommendations with MyBehaviorCBP, participants reported that they liked the personalization of MyBehaviorCBP. They also mentioned MyBehaviorCBP recommendations were more actionable and easier, and they were more likely to succeed if they tried the recommendations.

> *I really liked the personalization. I thought it was a nice touch. Suggestions were more specific and tailored, which for me made them more relevant and likely for me to use them.* [P1]
>
> *...most of the suggestions were fairly easy; at least the ones that involved walking.* [P3]
>
> *Because the suggestions of MyBehaviorCBP were based on my own chosen activities, I was much more likely to follow them.* [P4]
>
> *I liked them more because it seemed more likely that I could do them—I was more likely, in my mind, to succeed.* [P6]
>
> *Again, because the suggestions were based on my activities, they felt more feasible. I didn't have to take the extra step of thinking about how I might get the right tools (eg, bike) or where I can do the suggested exercise.* [P9]
>
> *Other than changing the sitting behavior, I liked the fact that they seemed more do-able.* [P10]

In addition, some participants liked the specificity of the recommendations and how they could be carried out in a smaller piecewise manner.

> *...they were location specific, smaller chunks of time.* [P3]
>
> *...more detailed explanations/suggestions, based on past exercises logged, and having the location helped, too!* [P8]

Regarding the control phase suggestions, some participants struggled with their nonpersonalized nature and how they needed to plan ahead to execute them.

> *...which I wanted to do the longer suggestions in version 2 [ie, control phase], unless I scheduled or planned it, I couldn't do most of them.* [P8]
>
> *...I received the suggestion to ride a bike, but that's currently simply not possible, logistically.* [P1]

However, one participant did not like MyBehaviorCBP recommendations and wanted more variety.

> *There was very little variation in the suggestions during the final 2 weeks—almost everyday was walk slightly farther and play tennis for a few minutes more... In the 2 to 3 weeks, there was a greater variety of things to try and I tried a few novel suggestions.* [P2]

Participants also asked for the following features: (1) a reminder system to plan in the morning and notifications in the moment, (2) adapt suggestions based on weather or weekend/weekday, and (3) better insight to relate high pain days and activity level, etc.

> *It would be helpful to have reminders and suggestions pop up in the morning or at other chosen times. This could be optional and set by the user.* [P1]
>
> *Maybe adding an alarm or something, to say "here, you should go do this thing now." I think if I had something bugging me to get up and take a short walk, for example, I would be more likely to do it than just looking at a list of things I might do.* [P3]
>
> *If it could ask me to rank the things I enjoy doing and then download weather data for the following days. This could suggest times when I have performed these tasks in the past and also match it with weather predictions. "You played tennis last Tuesday in the afternoon for 90 minutes. How about from 2 to 4 today when the weather will be clear and 85."* [P7]
>
> *Maybe a tally at the end of each week regarding days unable to exercise, based on back pain.* [P10]

Finally, one participant wanted to use the app even after the study, and mentioned the following:

> *I liked this app and look forward to possibly using it permanently in the future.* [P9]

Discussion

Principal Findings

To the best of our knowledge, MyBehaviorCBP is the first mobile app to provide automatically generated data driven physical activity recommendations in the chronic pain context. We conducted a pilot study to examine the feasibility of the approach. In the study, we found participants used the MyBehaviorCBP app 1 or more times a day. Furthermore, we observed early indication of acceptance and efficacy in both the qualitative and quantitative data. For instance, in the daily surveys, participants perceived the tracked data-based recommendations to be easier to follow. In the qualitative feedback on the exit survey, participants were positive to successfully complete MyBehaviorCBP recommendations. This means participants likely had a greater sense of self-efficacy toward MyBehaviorCBP-generated suggestions. According to protection motivation theory, higher self-efficacy may cause the recommendations to be carried out despite the presence of fear in chronic pain [58]. Indeed, empirically the participants also acted on the MyBehaviorCBP suggestions according to the automated walking inference. This increase gives early evidence that the perceived easiness may have transferred to actual behavior. However, the change in nonwalking exercise was not significant. This may have happened because it is hard to ascertain regularity in exercising within a 2-week period [35]. Furthermore, there was a small reduction of reported pain, which was not significant. This can be a type II error, and the small-scale study may not be sufficient to reject the system's efficacy [54].

Scopes for Future Improvement

From Figure 4, we observed that the number of times participants opened the app decreased over time. This may mean engagement decreased over time. Fortunately, participants also gave us valuable feedback to improve future versions of the app, which may increase engagement. For instance, several of our participants asked for notifications when the context was appropriate for a recommendation. Indeed, Fogg [39] argues that a trigger or notification may be necessary even when the suggested action can be executed with less effort. However, providing just-in-time notifications has particular technical barriers since it requires constant monitoring to detect the right context without draining the phone battery. In addition, the notifications can interrupt the participants' daily workflow, and it is unclear when the right time to provide an intervention would be. Therefore, future research may focus on just-in-time interventions that are acceptable to participants and more battery efficient.

Relationship With Earlier MyBehavior Work

The current MyBehaviorCBP system is a variant of a prior system, MyBehavior [36]. MyBehaviorCBP borrows several ideas from MyBehavior, such as clustering routine activities and the MAB algorithm. However, there are a few key differences between MyBehavior and MyBehaviorCBP. MyBehavior was designed for a weight loss population with no chronic pain, and the objective function of the MAB was to maximize calorie expenditure. On the other hand, MyBehaviorCBP is designed for people with chronic pain, and the goal of the MAB is to maximize number of minutes performing low-effort exercises. As a result, between walking and gym exercise, MyBehaviorCBP would prioritize walking since walking is less effortful, but MyBehavior would prioritize gym exercises since gym exercises give more calorie expenditure. In addition, MyBehaviorCBP provides recommendations such that the total number of recommended minutes to exercise does not exceed 60 minutes. MyBehavior did not have any limit on total calories for the recommendations.

Despite the similarities between the two systems, the effect sizes in Table 2 are lower than those in the previous MyBehavior trials [36]. This is expected, since MyBehavior was designed for weight loss and did not deal with the negative psychological challenges in chronic pain. Indeed, out of the 321 reported daily surveys for the MyBehaviorCBP, in 39.9% cases negative emotions were reported in the PAM [57]. In a study with MyBehavior, out of 687 daily survey responses, 32.1% cases negative emotions were reported. Therefore, MyBehaviorCBP likely made participants active despite the prevalence of negative emotions.

Relationship With Prior mHealth Apps for Chronic Pain Self-Management

Over the years, a variety of mobile apps have been proposed for chronic pain self-management, with some apps aiming at prescribing cognitive behavioral components [15,30,32]. Other apps use diary-based approaches by logging subjective self-reports and provide basic feedback such as reminders for medication [12], and some newer apps such as the WebMD PainCoach [31] enable pain monitoring, setting and tracking activity goals, and generating related messages. However, MyBehaviorCBP differs in that it uses an in-phone machine learning approach directly on tracked physical activity data and automatically generates new person-specific suggestions based entirely on this. This also lends MyBehaviorCBP's automated approach to be complementary to other approaches (eg, methods based on cognitive behavioral therapy that could be combined with the tracked data-driven suggestions from MyBehaviorCBP).

Limitations

One limitation is the small number of study participants and relatively short study length. However, MyBehaviorCBP is an early stage technology. It is difficult to acquire resources to conduct efficacy trials with unproven technology on a potentially vulnerable chronic pain population. As a result, the purpose of this pilot study was to inform feasibility and acceptability. In our future work, we will use the lessons learned in this pilot study to conduct longer term studies on a larger population and also to include specific back pain outcomes such as the Oswestry Low Back Pain Disability Questionnaire [59].

Another limitation is that MyBehaviorCBP does not fully address the question of whether even moderate exercise can have adverse consequences. If there is a short-term pain flare, then moderate exercising can temporarily increase pain and MyBehaviorCBP should not recommend exercising during a pain flare. However, it is not clear whether exercising has any long-term adverse effect on pain. Some prior work [60,61] argues exercising may have longer term effects and suggests activity pacing, where a participant engages in an active phase of exercise for a certain amount of days followed by a resting phase with decreased or no exercise. However, we are not aware of any published studies that have examined the question of long-term adverse effects empirically. Some providers recommend rest while others recommend that patients remain active without "overdoing it" based on their clinical experience. Future research is needed to address this important question of longer term effects. Using apps like MyBehaviorCBP could help address this question. For example, the static upper limit of a 60 minutes threshold could be dynamic. This value could be determined either automatically by a model-based approach or by an expert such as a physiotherapist. Should it be deemed that the user needs rest, the limit could be set to zero or a low value.

Conclusion

In this paper, we presented the acceptability and feasibility of MyBehaviorCBP, a data-driven physical activity recommender system for chronic pain. We found preliminary evidence of increased walking activity; a few key areas of improvements have been also identified. In future work, we will incorporate these improvements and run a randomized controlled trial. If efficacy is demonstrated, then a technology like MyBehaviorCBP could have great promise because it is an automated system with no second person involved (eg, a physiotherapist). Also, all the data processing of MyBehaviorCBP is kept inside the phone which allows the app to preserve user privacy. Such automated and privacy-preserving features imply that MyBehaviorCBP has few barriers to scalability.

Acknowledgments

This project is funded by the Translational Research Institute for Pain in Later Life at Weill Cornell Medical College and the National Institute of Aging. MR is funded by the National Institute on Drug Abuse (NIDA P50 DA039838; PI: Linda Collins) and the National Institute on Alcohol Abuse and Alcoholism (NIAAA R01 AA023187; PI: Susan Murphy). We thank Elaine Wethington, Xiaoya Wu, Zhe Lin, and Minghao Li for their early input in this project.

Conflicts of Interest

TC co-founded and has equity interest in HealthRhythms Inc, which develops mobile phone–based systems for mental health. GG serves on the advisory committee of HealthRhythms. This company, however, does not have any commercial interest in the area of chronic pain.

References

1. Turk DC, Rudy TE. IASP taxonomy of chronic pain syndromes: preliminary assessment of reliability. Pain 1987 Aug;30(2):177-189. [Medline: 3670869]
2. Tracey I, Bushnell MC. How neuroimaging studies have challenged us to rethink: is chronic pain a disease? J Pain 2009 Nov;10(11):1113-1120. [doi: 10.1016/j.jpain.2009.09.001] [Medline: 19878862]

3. Bushnell MC, Ceko M, Low LA. Cognitive and emotional control of pain and its disruption in chronic pain. Nat Rev Neurosci 2013 Jul;14(7):502-511 [FREE Full text] [doi: 10.1038/nrn3516] [Medline: 23719569]
4. American Academy of Pain Medicine. Facts and Figures on Pain URL: http://www.painmed.org/patientcenter/facts_on_pain.aspx [accessed 2018-09-18] [WebCite Cache ID 72WRbJNds]
5. Gatchel RJ, Peng YB, Peters ML, Fuchs PN, Turk DC. The biopsychosocial approach to chronic pain: scientific advances and future directions. Psychol Bull 2007 Jul;133(4):581-624. [doi: 10.1037/0033-2909.133.4.581] [Medline: 17592957]
6. Stewart WF, Ricci JA, Chee E, Morganstein D. Lost productive work time costs from health conditions in the United States: results from the American Productivity Audit. J Occup Environ Med 2003 Dec;45(12):1234-1246. [doi: 10.1097/01.jom.0000099999.27348.78] [Medline: 14665809]
7. Martel MO, Wideman TH, Sullivan MJL. Patients who display protective pain behaviors are viewed as less likable, less dependable, and less likely to return to work. Pain 2012 Apr;153(4):843-849 [FREE Full text] [doi: 10.1016/j.pain.2012.01.007] [Medline: 22325743]
8. Cook AJ, Brawer PA, Vowles KE. The fear-avoidance model of chronic pain: validation and age analysis using structural equation modeling. Pain 2006 Apr;121(3):195-206. [doi: 10.1016/j.pain.2005.11.018] [Medline: 16495008]
9. McWilliams LA, Cox BJ, Enns MW. Mood and anxiety disorders associated with chronic pain: an examination in a nationally representative sample. Pain 2003 Nov;106(1-2):127-133. [Medline: 14581119]
10. Geneen LJ, Moore RA, Clarke C, Martin D, Colvin LA, Smith BH. Physical activity and exercise for chronic pain in adults: an overview of Cochrane Reviews. Cochrane Database Syst Rev 2017 Dec 24;4:CD011279 [FREE Full text] [doi: 10.1002/14651858.CD011279.pub3] [Medline: 28436583]
11. International Association for the Study of Pain. 2012. Unrelieved pain is a major global healthcare problem URL: http://www.iasp-pain.org/AM/Template.cfm?Section=Home&Template=/CM/ContentDisplay.cfm&ContentID=2908 [WebCite Cache ID 72F9OTH15]
12. Singh A, Klapper A, Jia J, Fidalgo A, Tajadura-Jimenez A, Kanakam N, et al. Motivating people with chronic pain to do physical activity: opportunities for technology design. 2014 Presented at: Proceedings of the 32nd annual ACM conference on Human factors in computing systems; 2014; Toronto p. 2803-2812. [doi: 10.1145/2556288.2557268]
13. Fricton J, Anderson K, Clavel A, Fricton R, Hathaway K, Kang W, et al. Preventing chronic pain: a human systems approach-results from a massive open online course. Glob Adv Health Med 2015 Sep;4(5):23-32 [FREE Full text] [doi: 10.7453/gahmj.2015.048] [Medline: 26421231]
14. Elzahaf RA, Tashani OA, Unsworth BA, Johnson MI. The prevalence of chronic pain with an analysis of countries with a Human Development Index less than 0.9: a systematic review without meta-analysis. Curr Med Res Opin 2012 Jul;28(7):1221-1229. [doi: 10.1185/03007995.2012.703132] [Medline: 22697274]
15. Centers for Disease Control and Prevention. Vital signs: overdoses of prescription opioid pain relievers—United States, 1999-2008 URL: https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6043a4.htm?s_cid=mm6043a4_w [accessed 2018-02-15] [WebCite Cache ID 6xGCMAAyC]
16. National Centers for Health Statistics. Chartbook on Trends in the Health of Americans: Health, United States, 2006 URL: https://www.cdc.gov/nchs/data/hus/hus06.pdf [accessed 2018-02-15] [WebCite Cache ID 6xGCQh2Hy]
17. Dahm KT, Brurberg KG, Jamtvedt G, Hagen KB. Advice to rest in bed versus advice to stay active for acute low-back pain and sciatica. Cochrane Database Syst Rev 2010 Jun 16(6):CD007612. [doi: 10.1002/14651858.CD007612.pub2] [Medline: 20556780]
18. Dowell D, Haegerich TM, Chou R. CDC guideline for prescribing opioids for chronic pain—United States, 2016. JAMA 2016 Apr 19;315(15):1624-1645. [doi: 10.1001/jama.2016.1464] [Medline: 26977696]
19. Chou R, Deyo R, Friedly J, Skelly A, Hashimoto R, Weimer M, et al. Nonpharmacologic therapies for low back pain: a systematic review for an American College of Physicians Clinical Practice Guideline. Ann Intern Med 2017 Apr 04;166(7):493-505. [doi: 10.7326/M16-2459] [Medline: 28192793]
20. Department of Veterans Affairs. 2017. VA/DoD clinical practice guideline for opioid therapy for chronic pain, version 3.0 URL: https://www.healthquality.va.gov/guidelines/Pain/cot/VADoDOTCPG022717.pdf [accessed 2018-09-18]
21. Turk DC, Okifuji A. Psychological factors in chronic pain: evolution and revolution. J Consult Clin Psychol 2002 Jun;70(3):678-690. [Medline: 12090376]
22. Asmundson G, Norton P, Vlaeyen J. Fear-avoidance models of chronic paan overview. In: Asmundson G, Vlaeyen J, Crombez G, editors. Understanding and Treating Fear of Pain. Oxford: Oxford University Press; 2004:3-24.
23. Vlaeyen JW, Kole-Snijders AM, Rotteveel AM, Ruesink R, Heuts PH. The role of fear of movement/(re)injury in pain disability. J Occup Rehabil 1995 Dec;5(4):235-252. [doi: 10.1007/BF02109988] [Medline: 24234727]
24. Sullivan MJ, Thorn B, Haythornthwaite JA, Keefe F, Martin M, Bradley LA, et al. Theoretical perspectives on the relation between catastrophizing and pain. Clin J Pain 2001 Mar;17(1):52-64. [Medline: 11289089]
25. Jensen MP, Romano JM, Turner JA, Good AB, Wald LH. Patient beliefs predict patient functioning: further support for a cognitive-behavioural model of chronic pain. Pain 1999 May;81(1-2):95-104. [Medline: 10353497]
26. Cipher DJ, Fernandez E. Expectancy variables predicting tolerance and avoidance of pain in chronic pain patients. Behav Res Ther 1997 May;35(5):437-444. [Medline: 9149453]

27. Lethem J, Slade PD, Troup JD, Bentley G. Outline of a Fear-Avoidance Model of exaggerated pain perception. Behav Res Ther 1983;21(4):401-408. [Medline: 6626110]
28. della Volpe R, Popa T, Ginanneschi F, Spidalieri R, Mazzocchio R, Rossi A. Changes in coordination of postural control during dynamic stance in chronic low back pain patients. Gait Posture 2006 Nov;24(3):349-355. [doi: 10.1016/j.gaitpost.2005.10.009] [Medline: 16311036]
29. Dorflinger L, Kerns R, Auerbach S. Providers' roles in enhancing patients' adherence to pain self management. Transl Behav Med 2013 Mar;3(1):39-46 [FREE Full text] [doi: 10.1007/s13142-012-0158-z] [Medline: 24073159]
30. Ehde DM, Dillworth TM, Turner JA. Cognitive-behavioral therapy for individuals with chronic pain: efficacy, innovations, and directions for research. Am Psychol 2014;69(2):153-166. [doi: 10.1037/a0035747] [Medline: 24547801]
31. WebMD PainCoach. 2016. URL: https://www.webmd.com/webmdapp [accessed 2018-02-15] [WebCite Cache ID 6xGCThEDe]
32. Rathbone AL, Clarry L, Prescott J. Assessing the efficacy of mobile health apps using the basic principles of cognitive behavioral therapy: systematic review. J Med Internet Res 2017 Nov 28;19(11):e399 [FREE Full text] [doi: 10.2196/jmir.8598] [Medline: 29187342]
33. Lane N, Miluzzo H, Lu H, Peebles D, Choudhury T, Campbell A. A survey of mobile phone sensing. IEEE Comm 2010;48(9):140-150. [doi: 10.1109/MCOM.2010.5560598]
34. Aung S, Alquaddoomi F, Hsieh C, Rabbi M, Yang L, Pollak J. Leveraging multi-modal sensing for mobile health: a case review in chronic pain. IEEE J Sel Topics Sig Proc 2010 Sep;10(5):962-974. [doi: 10.1109/mcom.2010.5560598]
35. Rabbi M, Aung M, Zhang M, Choudhury T. MyBehavior: automatic personalized health feedback from user behaviors and preferences using smartphones. 2015 Presented at: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; 2015; New York p. 707-718. [doi: 10.1145/2750858.2805840]
36. 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]
37. Singh A, Piana S, Pollarolo D, Volpe G, Varni G, Tajadura-Jiménez A, et al. Go with the flow: tracking, analysis and sonification of movement and breathing to build confidence in activity despite chronic pain. Hum–Comput Interact 2015 Nov 25;31(3-4):335-383. [doi: 10.1080/07370024.2015.1085310]
38. Felipe A. Roles for personal informatics in chronic pain. 2015 Presented at: 9th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth); 2015; Istanbul p. 161-168. [doi: 10.4108/icst.pervasivehealth.2015.259501]
39. Fogg B. A behavior model for persuasive design. 2009 Presented at: Proceedings of the 4th international Conference on Persuasive Technology; 2009; Claremont. [doi: 10.1145/1541948.1541999]
40. Klasnja P. How to evaluate technologies for health behavior change in HCI research. 2011 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 2011; Vancouver p. 3063-3072. [doi: 10.1145/1978942.1979396]
41. Dallery J, Cassidy RN, Raiff BR. Single-case experimental designs to evaluate novel technology-based health interventions. J Med Internet Res 2013 Feb 08;15(2):e22 [FREE Full text] [doi: 10.2196/jmir.2227] [Medline: 23399668]
42. National Institutes of Health. 2011. Getting started and staying active URL: http://www.nhlbi.nih.gov/health/health-topics/topics/phys/getstarted [accessed 2015-05-07] [WebCite Cache ID 6YLlZqFNZ]
43. National Institutes of Health. 2013. Healthy eating plan URL: http://www.nhlbi.nih.gov/health/public/heart/obesity/lose_wt/calories.htm [accessed 2018-09-03] [WebCite Cache ID 6YEhYdWTL]
44. National Institutes of Health. Handout on health: back pain URL: https://www.niams.nih.gov/health-topics/back-pain [accessed 2018-02-16] [WebCite Cache ID 6xGWH3ENg]
45. Lu H. The Jigsaw continuous sensing engine for mobile phone applications. 2010 Presented at: Proceedings of the 8th ACM conference on embedded networked sensor systems; 2010; Zurich p. 71-84. [doi: 10.1145/1869983.1869992]
46. Ainsworth BE, Haskell WL, Herrmann SD, Meckes N, Bassett DR, Tudor-Locke C, et al. 2011 Compendium of Physical Activities: a second update of codes and MET values. Med Sci Sports Exerc 2011 Aug;43(8):1575-1581. [doi: 10.1249/MSS.0b013e31821ece12] [Medline: 21681120]
47. Tewari A, Murphy S. From ads to interventions: contextual bandits in mobile health. In: Mobile Health, pp. 495-517. Springer, Cham. New York: Springer; 2017:495-517.
48. Bertsekas D. Approximate Dynamic Programming. Hoboken: John Wiley & Sons; 2011.
49. Bandura A. Self-efficacy mechanism in human agency. Am Psychol 1982;37(2):122-147. [doi: 10.1037//0003-066x.37.2.122]
50. Badanidiyuru A. Bandits with knapsacks. 2013 Presented at: Foundations of Computer Science (FOCS), IEEE 54th Annual Symposium; 2013; Washington p. 207-216. [doi: 10.1109/focs.2013.30]
51. Ahtinen A, Mattila E, Välkkynen P, Kaipainen K, Vanhala T, Ermes M, et al. Mobile mental wellness training for stress management: feasibility and design implications based on a one-month field study. JMIR Mhealth Uhealth 2013;1(2):e11 [FREE Full text] [doi: 10.2196/mhealth.2596] [Medline: 25100683]

52. Klasnja P, Hekler EB, Shiffman S, Boruvka A, Almirall D, Tewari A, et al. Microrandomized trials: an experimental design for developing just-in-time adaptive interventions. Health Psychol 2015 Dec;34 Suppl:1220-1228. [doi: 10.1037/hea0000305] [Medline: 26651463]
53. Singer JD, Willett JB. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. Oxford: Oxford University Press; 2003.
54. Ott RL, Longnecker MT. An Introduction to Statistical Methods and Data Analysis. Scarborough: Nelson Education; 2015.
55. 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;13(3):e55 [FREE Full text] [doi: 10.2196/jmir.1838] [Medline: 21840837]
56. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77. [doi: 10.1007/978-1-4614-5583-7_311]
57. Pollak J. PAM: a photographic affect meter for frequent, in situ measurement of affect. 2011 Presented at: Proceedings of the SIGCHI conference on Human factors in computing systems; 2011; Vancouver p. 725-734. [doi: 10.1145/1978942.1979047]
58. Boer H, Seydel E. Protection Motivation Theory. Buckingham: Open University Press; 1996.
59. Davidson M, Keating JL. A comparison of five low back disability questionnaires: reliability and responsiveness. Phys Ther 2002 Jan;82(1):8-24. [Medline: 11784274]
60. Keefe FJ, Ahles TA, Sutton L, Dalton J, Baucom D, Pope MS, et al. Partner-guided cancer pain management at the end of life: a preliminary study. J Pain Symptom Manage 2005 Mar;29(3):263-272. [doi: 10.1016/j.jpainsymman.2004.06.014] [Medline: 15781177]
61. Gill JR, Brown CA. A structured review of the evidence for pacing as a chronic pain intervention. Eur J Pain 2009 Feb;13(2):214-216 [FREE Full text] [doi: 10.1016/j.ejpain.2008.03.011] [Medline: 18448368]

Abbreviations

AIC: Akaike information criterion
BIC: Bayesian information criterion
LR: likelihood ratio
MAB: multi-armed bandit
MET: metabolic equivalent of task
mHealth: mobile health
PAM: Photographic Affect Meter
RL: reinforcement learning

Edited by G Wadley, R Calvo, M Czerwinski, J Torous; submitted 15.02.18; peer-reviewed by E Heckler, S Kardeş; comments to author 25.03.18; revised version received 12.05.18; accepted 19.06.18; published 26.10.18.

Please cite as:
Rabbi M, Aung MSH, Gay G, Reid MC, Choudhury T
Feasibility and Acceptability of Mobile Phone–Based Auto-Personalized Physical Activity Recommendations for Chronic Pain Self-Management: Pilot Study on Adults
J Med Internet Res 2018;20(10):e10147
URL: http://www.jmir.org/2018/10/e10147/
doi:10.2196/10147
PMID:30368433

Letter to the Editor

Blood Pressure Measurement Validation Off the Cuff? Comment on “A New Cuffless Device for Measuring Blood Pressure: A Real-Life Validation Study”

Noud van Helmond[1], MD; Timothy B Plante[2], MD, MHS; Jeffrey I Joseph[1], DO

[1]Department of Anesthesiology, Sidney Kimmel Medical College, Thomas Jefferson University, Philadelphia, PA, United States
[2]Department of Medicine, Larner College of Medicine, University of Vermont, Burlington, VT, United States

Corresponding Author:
Noud van Helmond, MD
Department of Anesthesiology
Sidney Kimmel Medical College
Thomas Jefferson University
1020 Locust Street
Philadelphia, PA, 19107-6731
United States
Phone: 1 7168809732
Email: noud.van.helmond@jefferson.edu

Related Articles:

Comment on: https://www.jmir.org/2016/5/e85/

Comment in: http://www.jmir.org/2018/10/e12477/

(J Med Internet Res 2018;20(10):e10089) doi:10.2196/10089

KEYWORDS
cuffless blood pressure monitor; blood pressure monitor validation; hypertension

We read with interest the validation study of the commercially available CheckMe blood pressure (BP) monitor by Schoot et al [1]. We agree with the authors that the development of cuffless BP devices may improve rates of out-of-office BP measurement for adults with hypertension and that the value of such devices depends on the devices’ accuracy. We feel there are aspects of their report where clarification might be helpful.

One of the reported strengths of the study is that it met European Society of Hypertension International Protocol (ESH-IP) validation guidelines [2]. However, some aspects of the study deviated from the ESH-IP, including obtaining of measurements in varying anatomical positions. Usually, validation is performed in the manufacturer’s recommended position of measurement, which is typically in the same position as calibration [3]. Additional information would be helpful to better understand the performance characteristics from the device itself and to assess for bias introduced from protocol deviations.

First, presenting correlation coefficients and scatterplots can help readers assess the relationship between measurements from the devices. Next, the authors present a relative difference in mean systolic BPs, which is near zero. Presenting mean absolute difference between devices would be more informative to the accuracy of the device [2]. A modified version of the ESH-IP’s validation table is presented not in the usual numerical categories. Calculating these categories, the CheckMe deviates from the reference device by ≤5 mmHg, ≤10 mmHg, and ≤15 mmHg for 16%, 62%, and 86% of the measures, respectively. These performance characteristics do not meet the threshold of passing ESH-IP’s first part of validation. Finally, contrary to the description in the manuscript, we note that the reference device used has not undergone independent validation and its accuracy is not known [4]. Thus, it is difficult to interpret the accuracy of CheckMe without understanding the performance characteristics of the reference device.

We recognize that it is not possible to adhere to the ESH-IP with novel cuffless BP-measuring devices, since the ESH-IP assumes that the tested monitor does not require user-specific calibration. In response to the ongoing interest in cuffless BP measurement the Institute of Electrical and Electronics Engineers (IEEE) released a guideline specifically for the validation of these devices [5]. Because validation of a cuffless BP monitor immediately after calibration at the same BP could artificially increase its perceived accuracy [3], the IEEE protocol includes validation measurements after changes in BP from the calibration measurement and validation measurements after a significant period since calibration.

In conclusion, we encourage the authors to present additional analyses and results to improve the understanding of the CheckMe's performance, and to follow up with a formal IEEE-protocol validation study. In the present study, the device is reported to be a Conformité Européenne-approved category IIa category medical device compliant with directive 93/42/European Economic Community, but it did not meet that directive's accuracy requirements (EN 1060-3). The CheckMe has United States Food & Drug Administration approval for measurements of heart rate, oxygen saturation, temperature, and activity, but not for measurement of systolic blood pressure [6]. As the device does not meet ESH-IP accuracy criteria for blood pressure in this study, we are concerned that it is available for sale directly to consumers and may place adults at undue harm.

Conflicts of Interest

None declared.

References

1. Schoot TS, Weenk M, van de Belt TH, Engelen LJLPG, van Goor H, Bredie SJH. A New Cuffless Device for Measuring Blood Pressure: A Real-Life Validation Study. J Med Internet Res 2016 May 05;18(5):e85 [FREE Full text] [doi: 10.2196/jmir.5414] [Medline: 27150527]
2. O'Brien E, Atkins N, Stergiou G, Karpettas N, Parati G, Asmar R, Working Group on Blood Pressure Monitoring of the European Society of Hypertension. European Society of Hypertension International Protocol revision 2010 for the validation of blood pressure measuring devices in adults. Blood Press Monit 2010 Feb;15(1):23-38. [doi: 10.1097/MBP.0b013e3283360e98] [Medline: 20110786]
3. Boubouchairopoulou N, Kollias A, Chiu B, Chen B, Lagou S, Anestis P, et al. A novel cuffless device for self-measurement of blood pressure: concept, performance and clinical validation. J Hum Hypertens 2017 Jul;31(7):479-482. [doi: 10.1038/jhh.2016.101] [Medline: 28124684]
4. Medaval. 2017. Welch Allyn Vital Signs 300 Series URL: http://medaval.ie/device/welch-allyn-vital-signs-300-series/ [accessed 2018-02-03] [WebCite Cache ID 6wxcESAO4]
5. Institute of Electrical and Electronics Engineers. 2014 Aug 26. IEEE 1708-2014 Standard for Wearable Cuffless Blood Pressure Measuring Devices URL: https://standards.ieee.org/standard/1708-2014.html [accessed 2018-09-26] [WebCite Cache ID 72jAQRSZx]
6. U.S. Department of Health and Human Services Food and Drug Administration, Checkme Pro Health Monitor. 2015 Dec 10. URL: https://www.accessdata.fda.gov/cdrh_docs/pdf15/K150869.pdf [accessed 2018-02-08] [WebCite Cache ID 6x4IH3xMC]

Edited by G Eysenbach; submitted 09.02.18; peer-reviewed by J Goris; comments to author 26.09.18; revised version received 27.09.18; accepted 10.10.18; published 22.10.18.

Please cite as:
van Helmond N, Plante TB, Joseph JI
Blood Pressure Measurement Validation Off the Cuff? Comment on “A New Cuffless Device for Measuring Blood Pressure: A Real-Life Validation Study”
J Med Internet Res 2018;20(10):e10089
URL: http://www.jmir.org/2018/10/e10089/
doi:10.2196/10089
PMID:30348627

Letter to the Editor

Authors’ Reply: Comment on “A New Cuffless Device for Measuring Blood Pressure: A Real-Life Validation Study”

Sebastian Bredie[1], MD, PhD; Tom van de Belt[2], PhD; Harry van Goor[3], MD, PhD, FRCS

[1]Internal Medicine, Radboud University Medical Center, Nijmegen, Netherlands

[2]REshape Center for Innovation, Radboud University Medical Center, Nijmegen, Netherlands

[3]Surgery, Radboud University Medical Center, Nijmegen, Netherlands

Corresponding Author:
Sebastian Bredie, MD, PhD
Radboud University Medical Center
Internal Medicine
NIG 463
Nijmegen, 6500 HB
Netherlands
Phone: 31 243618819
Email: bas.bredie@radboudumc.nl

Related Articles:

Comment on: https://www.jmir.org/2016/5/e85/

Comment on: http://www.jmir.org/2018/10/e10089/

(J Med Internet Res 2018;20(10):e12477) doi:10.2196/12477

We thank van Helmond et al [1] for their critical appraisal of our recent study [2] on the use of the CheckMe, a cuffless blood pressure measuring device. Patient-friendly and easy-to-use blood pressure monitors will make an important contribution to self management of the cardiovascular risk profile. It is however important that we learn how these new devices perform in terms of reliability, accuracy, and to what extent they are usable in current practice of hypertension management.

For this reason, we carried out a very first real life validity study, in which we have followed the conditions of the European Society of Hypertension International Protocol (ESH-IP) as much as possible. This rightly raises a number of questions, which we have also addressed in the article. First of all, cuffless blood pressure measurement involves a completely different technique, using of the pulse transition time measured with a pulse oximeter and the electrical heart signal, measured with electrodes. By calibrating once with a classic bracelet blood pressure monitor, an estimate of the missing vessel wall compliance is made, which is then used as a constant factor for follow-up measurements with the CheckMe. Compared to classic Korotkoff tones or oscillometric measurement, new variables are introduced with the cuffless technique.

At present, a description in the ESH-IP is lacking of the way in which such calibration measurement, necessary for the cuffless blood pressure measurement technique, is to be performed. Therefore, no formal validation study can be performed and it is also not possible to determine whether the CheckMe formally meets the criteria stated in the ESH-IP, as we stated in the article.

In anticipation of expected developments in technology and protocoling, we performed a first 'real life' comparison, in which we performed the calibration measurement and gold standard blood pressure measurement with a frequently used automatic office blood pressure monitor. Outside the ESH-IP we examined whether the position of the CheckMe with respect to the heart, which changed the pulse transition time from heart to fingertip, had an effect on the results. We believe that pioneering work in the field of new techniques for e.g. blood pressure measurement stimulates further development. We fully agree that a formal assessment of accuracy must be made according to the consensus in international guidelines. However, in the absence of in this case the requirement for performing the calibration measurement, only that part of the protocol that applies can be followed.

We appreciate the suggestions made by Helmond et al to also present data in an alternative manner and seriously consider a follow-up manuscript for adequate presentation and interpretation. We also recognize that the blood pressure measurement is not included in the FDA approval of the CheckMe. The required pulse oximetry and ECG recording are included, but as we estimate, for the above reason an FDA approval for blood pressure has never been issued.

We appreciate that our preliminary scientific data on the cuffless blood pressure measurement in persons with a broad range of blood pressure values is correctly assessed. We hope this will trigger a discussion about the use of this promising technique in patient care and will follow formal validation studies based on an adapted protocol [3].

Conflicts of Interest

None declared.

References

1. van Helmond N, Plante TB, Joseph JI. Blood Pressure Measurement Validation Off the Cuff? Comment on “A New Cuffless Device for Measuring Blood Pressure: A Real-Life Validation Study”. J Med Internet Res 2018 Oct 19;20(10):e10089 [FREE Full text] [doi: 10.2196/10089]
2. Schoot TS, Weenk M, van de Belt TH, Engelen LJLPG, van Goor H, Bredie SJH. A New Cuffless Device for Measuring Blood Pressure: A Real-Life Validation Study. J Med Internet Res 2016 May 05;18(5):e85 [FREE Full text] [doi: 10.2196/jmir.5414] [Medline: 27150527]
3. Watanabe N, Bando YK, Kawachi T, Yamakita H, Futatsuyama K, Honda Y, et al. Development and Validation of a Novel Cuff-Less Blood Pressure Monitoring Device. JACC Basic Transl Sci 2017 Dec;2(6):631-642 [FREE Full text] [doi: 10.1016/j.jacbts.2017.07.015] [Medline: 30062178]

Edited by G Eysenbach; submitted 11.10.18; this is a non–peer-reviewed article;accepted 12.10.18; published 22.10.18.

Please cite as:
Bredie S, van de Belt T, van Goor H
Authors’ Reply: Comment on “A New Cuffless Device for Measuring Blood Pressure: A Real-Life Validation Study”
J Med Internet Res 2018;20(10):e12477
URL: http://www.jmir.org/2018/10/e12477/
doi:10.2196/12477
PMID:30348625

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org