JMIR Human Factors

Making health care interventions and technologies usable, safe, and effective
Volume 2 (2015), Issue 1 ISSN: 2292-9495

Contents

Original Papers

Review

Original Paper

Development and Pilot Evaluation of an Online Relapse-Prevention Program Based on Acceptance and Commitment Therapy for Chronic Pain Patients

Martine Fledderus[1,2], PhD; Karlein MG Schreurs[1,2*], PhD; Ernst T Bohlmeijer[2*], PhD; Miriam MR Vollenbroek-Hutten[1*], PhD

[1]Roessingh Research and Development, Enschede, Netherlands

[2]Faculty of Behavioural Sciences, Department of Psychology, Health and Technology, University of Twente, Enschede, Netherlands

[*]these authors contributed equally

Corresponding Author:
Martine Fledderus, PhD
Roessingh Research and Development
Roessinghsbleekweg 33b
Enschede, 7500 AH
Netherlands
Phone: 31 534894470
Fax: 31 534892388
Email: martinefledderus@gmail.com

Abstract

Background: A significant number of chronic pain patients experience a decline in therapeutic effects after rehabilitation. As face-to-face contacts with health care professionals are not always feasible after treatment, new, innovative, fully automated relapse-prevention programs are highly needed.

Objective: In this study an online, automated relapse-prevention program based on acceptance and commitment therapy (ACT)—both as a website and as a mobile app—was developed and evaluated. At each step of the development, end users (ie, chronic pain patients) were consulted in order to fully address their needs.

Methods: In a step-by-step process, a contextual inquiry, requirement specification, and design were executed with chronic pain patients by conducting, respectively, a focus group (n=10), interviews with rapid prototyping (n=28), and a user- and expert-based usability evaluation (n=14). Furthermore, a pilot evaluation was conducted with 14 chronic pain or fatigue patients who had received the online relapse-prevention program following a multidisciplinary ACT treatment. They were interviewed about their usage and the usefulness of the program in supporting them to maintain changed behaviors and prevent relapses in avoidance and pain control behaviors.

Results: The three stages provided information about the expected needs of end users, comments about the usefulness of the proposed features, and feedback about the design and usability of the program. This resulted in a fully operational, online relapse-prevention program. Results from the pilot evaluation showed that 9 patients used the online program after treatment, 5 of whom indicated that the program supported them after treatment. Of all the patients, 4 of them indicated that the program did not support them because they wanted more social interaction with other users.

Conclusions: This study showed that an innovative, automated, online program that is user friendly can be developed by involving the end users in each step. The program was evaluated positively by some participants. The evaluation showed that the online relapse-prevention program has the potential to support chronic pain patients in maintaining their changed behaviors and preventing relapses in avoidance and pain control behaviors.

Trial Registration: Nederlands Trial Register (NTR) Number: NTR4177; http://www.trialregister.nl/trialreg/admin/rctview.asp?TC=4177 (Archived by WebCite at http://www.webcitation.org/6Ur6EFD1D).

(JMIR Human Factors 2015;2(1):e1) doi:10.2196/humanfactors.3302

KEYWORDS

chronic pain; eHealth; acceptance and commitment therapy; relapse prevention; aftercare

Introduction

Multidisciplinary rehabilitation programs based on cognitive behavioral therapy (CBT) or acceptance and commitment therapy (ACT) for chronic pain patients have shown positive effects on the interference of pain in daily life and on physical and mental functioning [1-3]. However, a significant number of patients experience a decrease in the therapeutic effects one year after rehabilitation [4,5]. Providing support after treatment might help to generate the skills required to prevent or manage the occurrence of a relapse. However, face-to-face contacts with a health care professional are not always feasible in rehabilitation care due to limited therapist time and a lack of (financial) resources [6]. A relapse-prevention program based on eHealth might overcome these barriers because it offers the user more convenience and more control over the content and timing of the intervention [7]. Moreover, it might be more cost-effective than face-to-face treatment as guidance can be given through email or short message service (SMS) text messaging [8]. A growing number of studies have shown that Web-based CBT interventions are effective for the treatment of chronic pain [9,10]. Two studies concluded that Web-based, CBT relapse-prevention programs following multidisciplinary pain treatment have shown positive effects [6,11].

In this study we describe the development of a new, online relapse-prevention program—in the form of a website and/or mobile app—based on ACT. The main focus of ACT is enhancing psychological flexibility which includes the processes of acceptance and value-based behavior [12]. For chronic pain patients, acceptance means that one acknowledges the pain and abandons unproductive attempts to control the pain [13]. When these attempts are relinquished, an individual can choose, or persist in, behaviors that are in line with life values [14]. Values are important, chosen life directions, for example, in the domains of family, work, and social life. Clients are encouraged to perform actions which are in line with their values, regardless of what emotions or thoughts might occur. Other important processes of ACT are mindfulness and self-as-context. These processes help a person to consciously center themselves in the present moment. This grounded awareness in the present moment is a necessary premise to be open and flexible to experience, and to move toward valued, day-to-day life activities [14]. The content of the online relapse-prevention program was based on cognitive behavior models of relapse, mostly applied in the area of substance abuse [15]. The most critical predictor of relapse is the individual's ability to perform effective coping strategies when facing high-risk situations. Therefore, relapses may be prevented by identification of these high-risk situations and teaching effective coping strategies. Other important determinants for preventing a relapse are high self-efficacy, functional social support, and positive affect [15].

However, despite the promising results of eHealth interventions for chronic pain, the effect sizes are often less than expected. One reason for this might lie in the low adherence rates of eHealth interventions, for instance, participants are not following the intervention as intended. For example, in the study of Moessner et al it was shown that a Web-based aftercare intervention following multidisciplinary therapy for chronic back pain had positive effects on pain intensity, but 70% of the patients did not adhere to the intervention [6]. Kristjánsdóttir et al only found a large effect size on the primary outcome (catastrophizing) in the participants who completed the mobile phone-based ACT intervention after a rehabilitation program for chronic widespread pain [16]. This corroborates earlier findings about the relationship between adherence and effect in psychological online interventions [17]. Studies on the underlying reasons for these low-adherence rates are scarce. However, commonly suggested reasons include shortcomings in the user-friendliness of the technology, problems with integrating the technology into day-to-day life, or a failure to tailor the intervention to the users' real needs [18].

The first aim of this study is to develop an innovative, fully automated, online relapse-prevention program that enables rehabilitation centers to implement this program. For the development process, a user-centered design was used that might overcome the problems described above. A user-centered design has been shown to have positive effects, especially on user-satisfaction levels and on addressing user needs [19]. A user-centered design gathers feedback from potential users during the whole development and design process [20]. The development and design processes were guided by a roadmap for creating an eHealth technology that follows the principles of a user-centered design [21]. The roadmap provides an overview of the different steps that need to be addressed with an explicit focus on involving all relevant stakeholders at each step for ensuring that the technology is broadly supported. Based on a review of current eHealth frameworks, the roadmap dictates six principles regarding technology development: (1) it is a participatory process, (2) it involves continuous evaluation cycles, (3) it is intertwined with implementation, (4) it changes the organization of health care, (5) it should involve persuasive technology, and (6) it needs advanced methods to assess impacts. Based on these principles, the roadmap consists of five research and development activities, which are contextual inquiry, requirement specification, design, operationalization, and summative evaluation [21]. This roadmap has been successfully used to develop an online, ACT, Web-based intervention for depression [22] and chronic pain [23]. In the current study, we initially focused on the first three processes to develop a new, innovative, online relapse-prevention program for chronic pain patients. Addressing the last two processes of the roadmap, the app is implemented in daily practice (operationalization), and the degree of successful uptake and the impact on health-related outcomes are evaluated (summative evaluation). To gain insights into these processes, we performed a pilot study where patients underwent multidisciplinary treatment. Following treatment, we evaluated whether patients found the program helpful in maintaining behavioral changes and preventing relapses in avoidance and pain control behaviors.

Methods

Contextual Inquiry

In the contextual inquiry, intended users are asked for information about their needs for a technology. Contextual inquiry also examines how the technology may fit into the

day-to-day life of the intended users [21]. In the current study, this information was obtained by a focus group discussion with chronic pain patients. A group of 10 female chronic pain patients who recently finished an 8-week, inpatient, multidisciplinary ACT program at a rehabilitation center in the Netherlands was invited to take part in the focus group discussion conducted by 2 researchers (KMGS, MF).

Respondents were asked to discuss what they believed would help them to prevent relapses and whether they would use a website and/or a mobile app as a relapse-prevention program. They were also asked whether they wanted to have guidance by email or SMS text message. This focus group session was audiotaped with the permission of the respondents. The audiotapes were transcribed and were analyzed qualitatively by the researcher by summarizing common themes.

Requirement Specification

In the requirement specification activity the expected needs are translated into requirements of the technology [21]. Therefore, based on the expected needs that were identified in the contextual inquiry, a prototype of a website page and various prototypes of pages of a mobile app were designed using PowerPoint. On multiple slides, five features of the online program were presented to the participants. For examples of these prototypes, see Figures 1 and 2. There were five features that demonstrated what a potential user could do in the online relapse-prevention program:

1. Values and actions: Users could add or change various life values and corresponding actions.

2. Diary: The diary is for monitoring value-based living and included one question that asked whether the user had lived according to his/her values on a rating scale from 1 to 10. The answers to this question were presented in a chart including positive smileys if a score of 6 or more was achieved.

3. Exercises: Users could add various ACT exercises or search in a database that contains all the exercises included in the treatment.

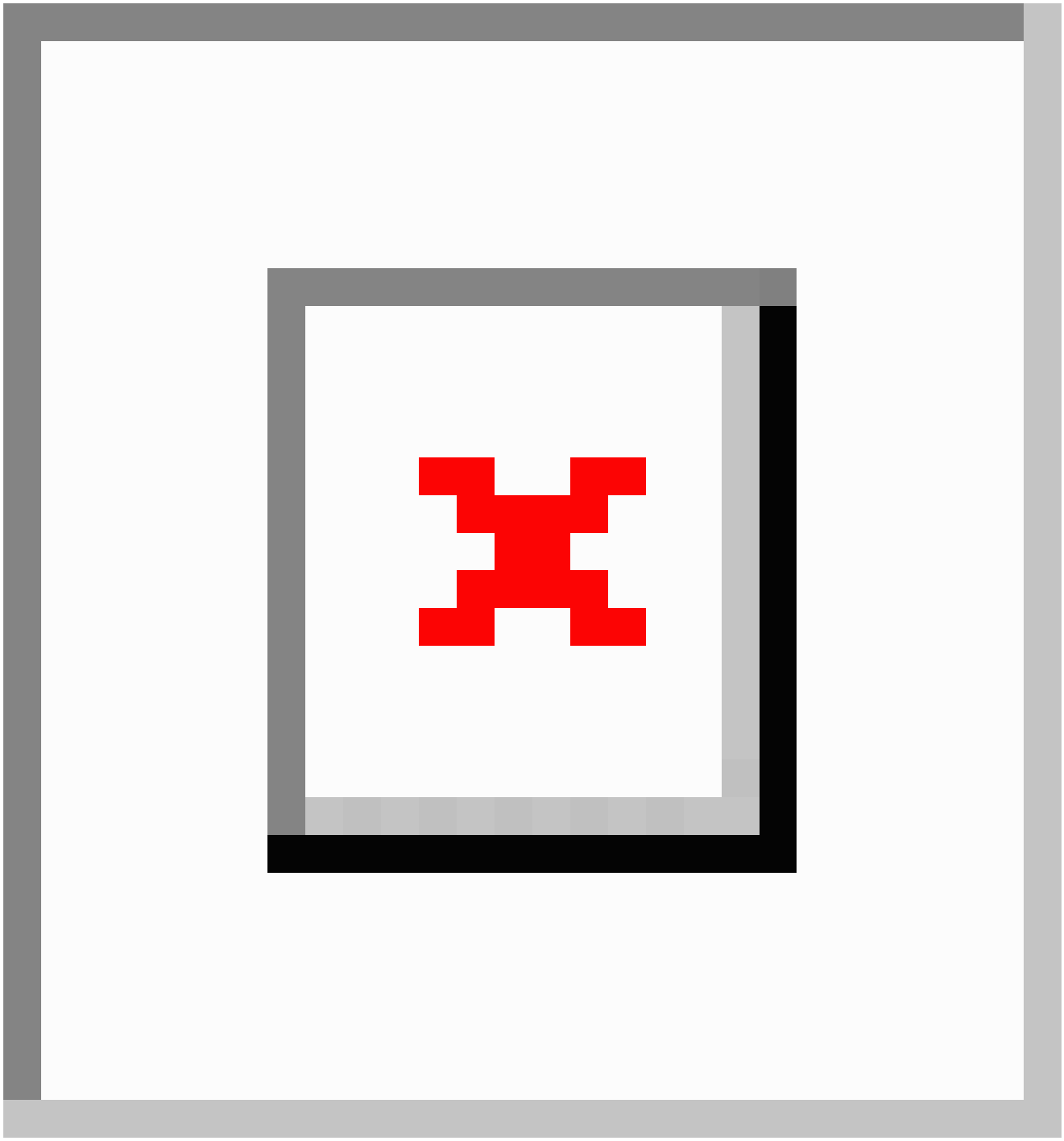
4. Tips: Users could add their own tips, share these tips with other users, or see shared tips from other users.

5. Coach: Various options of guidance with SMS text message or email were shown to the participants and they could all choose which text message they would like to receive: (1) a reminder SMS text message to use the diary or to use the program itself, (2) a motivational SMS text message written by a health care professional, (3) a self-composed motivational SMS text message which could be sent at a later date, or (4) a motivational SMS text message tailored to particular scores that are obtained from the diary—this tailors the content more to the individual situation.

To obtain the requirements, semistructured interviews combined with the prototypes of the eHealth technology were conducted with chronic pain patients [24]. Therefore, 28 chronic pain patients that recently completed the multidisciplinary ACT program at the rehabilitation center were interviewed by 3 psychology students from the University of Twente. Mean age was 43 years (SD 11) and most were female (25/28, 89%). The interview scheme was focused on the usefulness of the five different features of the online program by showing the prototypes based on these features. We asked the participants to comment on the design of these features by asking them to describe their first impression in three words and to indicate whether they found what was shown to them to be clear. If participants found the feature useful, they were asked to indicate the exact moment—during and/or after treatment—at which they would use these features. Finally, questions were asked about how they would like to receive the guidance (Coach). The audiotapes were transcribed and analyzed by the researcher using an inductive thematic analysis [25]. For this analysis the data were reviewed for identifying relevant patterns (themes) in the data. Initial codes were given to the responses of the participants and these codes were collated to potential themes. After this step the themes were again reviewed by checking the whole dataset and refined if necessary.

Figure 1. Example of a website prototype (Values and actions).

Figure 2. Example of a mobile app prototype (Exercises).



Design of the Technology

A fully operational program, both as a website and as a mobile app, was developed. For examples of these, see Figures 3 and 4. Based on the requirement specification, some changes were made. A library with examples of actions was added to the feature Values and actions. The name “Diary” was changed to “How are you?” and some extra tips written by a health care professional were added. For the Coach feature, several options for sending texts—both as email and SMS text message—were developed and users were free to choose between these options or to change them during use. Users can receive reminder messages after one week of not logging in or filling out the question in the “How are you?” section. The program includes an agenda for setting a fixed date and time to send the message. Users can receive a motivational message either once a week at random or within 24 hours after answering the question in the “How are you?” section. Furthermore, there was an option for the user to send self-composed motivational messages which could be sent once a week at random after treatment. All the reminder and motivational messages were written by 2 health care professionals (MF, KMGS). KMGS is a health care psychologist and is registered as a cognitive behavioral therapist with ample experience in ACT. She is a therapist in the multidisciplinary ACT team. MF has a PhD in psychology. Based on earlier knowledge, both in practice and earlier studies [23], they wrote all the messages in advance. The messages were programmed to be sent at random in the online program.

In this phase, the quality of the design was examined by evaluating its usability. Consequently, a user-based and expert-based evaluation method using a scenario-based think-aloud protocol was performed [24]. Usability tests with 5 chronic pain patients and 9 experts were conducted by 3 psychology students from the University of Twente. There were 5 target-group experts, such as a physiotherapist and psychologist who were working with chronic pain patients. Furthermore, there were 4 usability experts who were conducting research in eHealth. The mean age of the participants was 38 years (SD 12) and most of them were female (12/14, 86%). The participants were guided through the program using scenarios that included several tasks or problems that had to be solved by the user, such as "You want to add a new personal value to the program: how would you do that?" Participants were asked to verbalize their thoughts while they were taking part in these scenarios. The audiotapes were transcribed and the comments were defined as problems encountered and suggestions for improvements. A coding scheme based on system quality, content quality, and service quality was used [21,26]. System quality is defined as the user-friendliness of the program and the presentation of the content, such as the layout and where the buttons are placed. Content quality refers to the meaningfulness of the information. In other words, are the texts easy to understand and complete? Service quality refers to the degree to which the user assesses the service as being adequately provided, such as the perceived usefulness of the features and the provision of feedback [21,26]. Out of 14 participants, 8 of them (57%) evaluated the website and 6 of them (43%)—1 patient, 1 target-group expert, and 4 eHealth experts—evaluated the mobile app. Every participant could choose whether he/she wanted to evaluate the website or the mobile app. More eHealth experts evaluated the mobile app than patients. This might be explained by the fact that eHealth experts were more familiar with (mental health) apps at that moment. Every comment was evaluated as positive, negative, or neutral.

Figure 3. Screenshot of a page of the website (Values and actions).

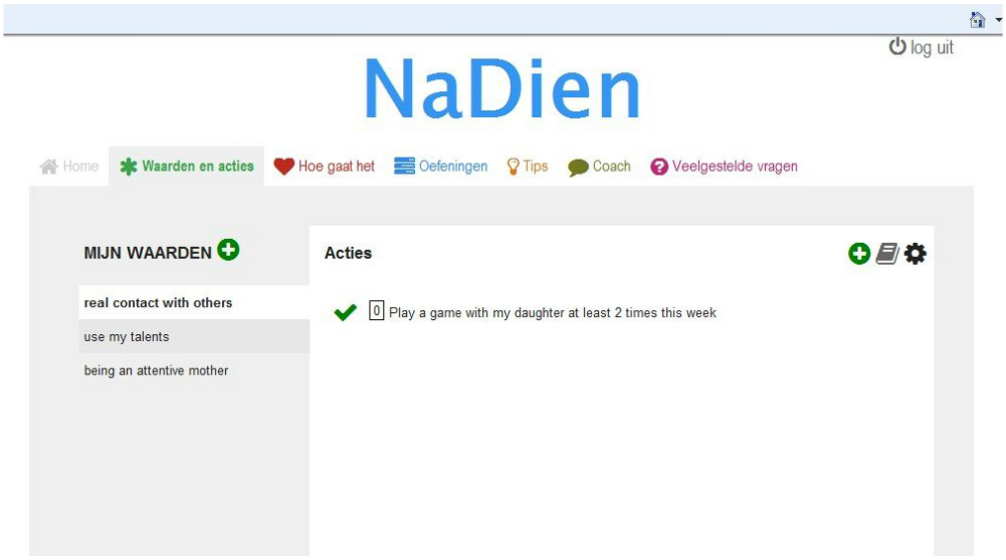
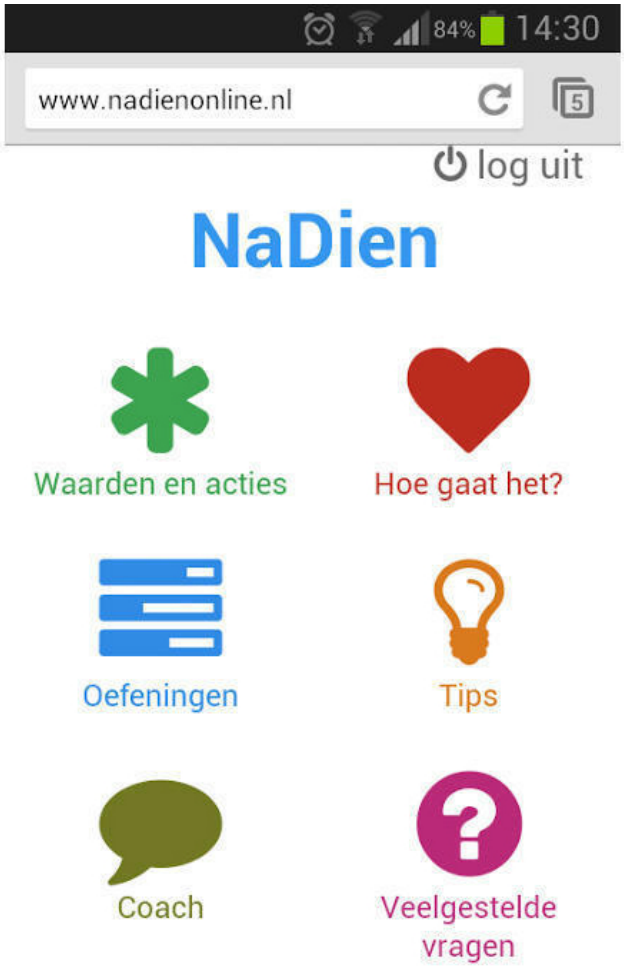


Figure 4. Screenshot of the mobile app homepage.



Pilot Evaluation of the Online Relapse-Prevention Program

Based on the results of the usability evaluation, the system quality was improved and a manual was developed to give the users more information, for example, whom the tips were shared with. A pilot user evaluation was conducted with patients who received the online program after a multidisciplinary pain treatment for 2 months. This study was approved by an independent Dutch medical ethics committee (Medical Research Ethics Committee Twente, no. P13-07) and recorded in the Dutch primary trial registry for clinical trials (Nederlands Trial Register, NTR4177).

Inclusion criteria were chronic pain or fatigue patients who were following an inpatient 8- to 12-week, group pain treatment which started in March or April 2013 at the Pain Department of the Roessingh Rehabilitation Centre. Chronic pain and chronic fatigue patients receive the same multidisciplinary treatment in the rehabilitation center. Patients needed to have access to the Internet through a computer and/or a mobile phone at home. The information letter described that it was recommended to use the online program for 5 to 10 minutes daily. The researcher had access to the actual accessed data use of the program, including how many times the participants were logged in and which features they had used, so self-report data use could be compared to the actual usage. In the fourth week of their treatment, the researcher explained the study and the patients received a letter with information about the study and a letter requesting their informed consent. In the sixth week of the treatment, patients who agreed to participate handed over the signed informed consent, were given access to the online ACT relapse-prevention program, and received a short introduction to the program. Following a period of 2 months after the group treatment, all of the participants were invited for a telephonic, structured interview conducted by the researcher (MF). The interviews were audiotaped with the permission of the participants and took, on average, 20 minutes to complete. The interview scheme started with a general question on the helpfulness of the online ACT relapse-prevention program in supporting them with the actions they intended to carry out after treatment. Participants were asked how many times they had used the program and whether they intended to continue using the program. The interview continued with an open discussion to identify which parts of the intervention they found to be most useful. Furthermore, they were also asked whether they would recommend this program to other patients. The interview ended with the question “Do you have any recommendations for improving the program?”

In total, there were 5 groups that started with the pain treatment with a total of 27 patients. All of them were given a letter with

information about the program and a form requesting their informed consent. The informed consent form was signed by 17 patients and they subsequently received the online program. The mean age of the participants was 38 years and most of them were female (14/17, 82%). All 17 patients were invited for the interview and 14 of them were interviewed by the researcher. The reasons for not taking part in the interview were medical (1/17, 6%), feeling very good (1/17, 6%), and unknown (1/17, 6%).

Results

Contextual Inquiry

All participants indicated that there is a need for an aftercare program and they would all like to have contact with their health professionals after treatment, for instance, through a telephone call or individual face-to-face contact. Most respondents would use an eHealth program, such as a website or mobile app. In this online program 8 participants out of 10 (80%) found it useful to register their values and actions and to read ACT-based exercises. Of the 10 participants, 4 of them (40%) thought it would be useful to make a start with completing these exercises and registering their values and actions during treatment. All participants would like to have contact with other patients to share tips and to give each other pep talks. Of the 10 participants, 2 of them (20%) proposed a forum for facilitating this contact with other patients. All participants would like to receive reminders by SMS text messaging or email from a health care professional to help them adhere to value-based behavior after treatment. Of the 10 participants, 8 of them (80%) indicated that reminders could be preformulated and sent automatically, preferably at a higher frequency rate and directly after the end of the treatment. Of the 10 participants, 3 of them (30%) also thought that it would be helpful to add their own motivational texts during treatment in the online program that can be sent after treatment. Additionally, 2 participants out of 10 (20%) indicated that the text messages should be tailored to the individual patient's personal situation.

Requirement Specification

Almost all of the participants indicated that they would use the online aftercare program (27/28, 96%). Some stated that they would like to use a website (12/28, 43%), a combination of a website and a mobile app (9/28, 32%), or only a mobile app (7/28, 25%). In general, participants found the design of the five features to be clear and convenient. Table 1 shows an outline of the participants' evaluation of the usefulness, the expected use, and needs of the various features of the online program. Almost all of the participants assessed all the features as useful and they indicated that they would use these features mostly during and after treatment. Furthermore, as can be seen in Table 1, the expected needs concerning the Coach feature are further specified and some improvements were given, such as changing the term "Diary".

Table 1. User evaluation of the features of the online relapse-prevention program (n=28).

Features	Usefulness, n (%)	Use, n (%)	Moment of use	Expected needs
Values and actions	28 (100%)	26 (93%)	During and after treatment.	Database with some examples of actions.
Diary	28 (100%)	25 (89%)	After treatment.	Changing the name "Diary".
Exercises	28 (100%)	27 (96%)	During and after treatment.	Database with some examples of exercises.
Tips	28 (100%)	27 (96%)	During and after treatment.	Tips from a health care professional.
Coach	28 (100%)	28 (100%)	After treatment.	Preference for receiving an SMS text message instead of an email.
				Receiving a reminder for the diary and technology.
				Choosing the frequency and moment to send out the reminders.
				Receiving an SMS text message with motivational content written by the health professional at random.
				Receiving an SMS text message with motivational content written by health professional after filling out the diary.
				Receiving a self-composed SMS text message.

Design of the Technology

All comments are represented in Table 2. They were largely the same between users and experts. The most positive comments were about the quality of the system. In particular, the ease of use of the website/app (eg, clear navigation, clear buttons, simple) and the design (eg, fresh, calm) were rated positively. The other most frequently mentioned positive comments were about the quality of the service, namely the perceived usefulness of the features (eg, adding personal values, sharing tips with other users, the options of receiving an SMS text message). The most negative comments were about the quality of the system as some technical errors occurred or icons were unclear. For example, the icon for sharing the tips was unclear because the participants did not recognize this icon. In the Diary feature, it was unclear that to register you had to slide the bar instead of clicking. The negative comments about the quality of the content were mostly linked to the Tips feature since the participants did not know with whom the tips were shared and, therefore, they did not want to use this feature.

Table 2. Number and positivity or negativity of comments yielded from the user- and expert-based methods.

Property of program	Number of comments by users[a]			Number of comments by experts[b]			Total
	+[c]	+/-[d]	-[e]	+	+/-	-	
System	45	5	22	52	10	65	199
Content	2	3	8	0	4	19	36
Service	15	0	3	11	2	10	41

[a]There were 5 users.

[b]There were 9 experts.

[c]Positive comment.

[d]Neutral comment.

[e]Negative comment.

Pilot Evaluation of the Online Relapse-Prevention Program

Overview

Table 3 shows the results of the pilot evaluation. There were 9 participants out of 14 (64%) who used the program and all of them will continue using the program. Out of 14, 5 participants (36%) never used the program at all, but most of them (3/5, 60%) indicated that they were going to use the program in the near future. Self-report data on the usage of the program were comparable to the actual accessed data usage. Almost all of the participants indicated that they would recommend the program to other patients (12/14, 86%). Furthermore, most participants (10/14, 71%) indicated that it would have been useful if they could have already started using the program during treatment, particularly with regard to completing certain parts of the program (eg, Values and actions, Tips), discussing it with other patients, and becoming better acquainted with it.

Table 3. Results of the pilot evaluation.

Participants (n=14)		n (%)
Nonusers		5 (36)
	Preference for use during treatment	4 (29)
	Recommend to other patients	5 (36)
	Use in future	3 (21)
	Most useful feature	
	Motivational messages	2 (14)
Users with positive evaluation [a]		5 (36)
	Preference for use during treatment	3 (21)
	Recommend to other patients	5 (36)
	Use in future	5 (36)
	Most useful feature	
	(Mindfulness) exercises	4 (29)
	Motivational messages	3 (21)
	Tips	3 (21)
Users with negative evaluation [b]		4 (29)
	Preference for use during treatment	3 (21)
	Recommend to other patients	2 (14)
	Use in future	4 (29)
	Most useful feature	
	Motivational messages	3 (21)
	Tips	1 (7)

[a]The online program was supportive in maintaining behavioral changes and preventing relapses in avoidance and pain control behaviors.

[b]The online program was not supportive in maintaining behavioral changes and preventing relapses in avoidance and pain control behaviors.

Nonusers

Reasons for nonuse were feeling well at that particular moment (3/14, 21%), a medical reason (1/14, 7%), and having other priorities (1/14, 7%). Of the 5 nonusers, 2 of them (40%) received motivational messages and found them very pleasant and supporting. Of the 5 nonusers, 3 of them (60%) thought they would use the program in the near future, especially in the case of a relapse (2/5, 40%). No suggestions for improvements were given.

Users With a Positive Evaluation

Out of the 14 participants, 5 of them (36%) indicated that the online program supported them in their efforts to perform their intended actions. Out of these 5 participants, 2 of them (40%) used the online program regularly (eg, at least once a week) and the other 3 participants (60%) used it 3 to 5 times in total. The online program functioned as a summary of the treatment and reminded them of what they had learned during treatment. The (mindfulness) exercises, the motivational messages, and the tips were evaluated as being "useful" and "very pleasant" as they functioned as reminders about the treatment. One participant recommended changing the question "How are you?" to more personal actions.

Users With a Negative Evaluation

Of the 14 participants, 4 of them (29%) indicated that the program did not support them after treatment. Of these 4 participants, 3 of them (75%) used the program 1 to 3 times, and 1 of them (25%) used the program at least once a week. All 4 participants out of 14 (29%) indicated that they missed the interaction with other patients in the program, for example, using a forum to share experiences with other users. Participants were satisfied with the motivational messages and the exercises because they served as a reminder for the treatment. They will continue to use the program, particularly the motivational messages (3/4, 75%), as well as the tips and the values and actions (1/4, 25%).

Discussion

Principal Findings

In this study, a new, innovative, online relapse-prevention program for chronic pain patients based on acceptance and commitment therapy was developed and evaluated. The first aim of the study was to develop an automated, user-friendly, online relapse-prevention program that fulfills the needs of chronic pain patients. A contextual inquiry by focus group discussion, requirement specification by interviews with rapid prototyping, and a user- and expert-based usability evaluation of the fully operational program successively provided the input for the next step in the development process. The chronic pain patients reviewed the program from their points of view and context, both on design and content for ensuring that the eHealth program was usable and acceptable. Accordingly, our user-centered developmental process resulted in a program with a simple design, large icons, and few layers of information and text. A pilot evaluation with chronic pain or fatigued patients who had received the online relapse-prevention program following a multidisciplinary ACT treatment (n=14) showed that this development process was satisfactory. Two-thirds of the participants (9/14, 64%) used the program in the 2 months after treatment. They did not have new suggestions on the usability of the program and nonusers (5/14, 36%) stated that their nonuse was not caused by complexity or inadequate usability of the program. This probably improves the uptake of the online relapse-prevention program, as many problems with adherence to eHealth programs are due to complexity and inadequate usability [18].

Furthermore, the user-centered development process resulted in information on the needs concerning the content of the program. The resulting program consists of the essential building blocks of a relapse-prevention program [15]. Firstly, participants have to recognize situations with a high risk of relapse. In ACT, relapse is defined as falling back on pain-avoidance behaviors instead of performing values-based actions [13]. By registering their values and actions and by monitoring values-based behavior with the tool "How are you?" participants are reminded of their important values and actions and recognize when they are relapsing to pain-avoidance behaviors. Next, patients have to have adequate coping skills to face high-risk situations. In the program, participants have access to descriptions of all exercises from which they learned these coping skills, their own favorite exercises, and tips from fellow users. Finally, social support after treatment was evaluated as an important component for the online relapse-prevention program. Social support can provide relevant shared advice and facilitates the process of finding recognition [27]. The program provides social support by offering the opportunity of sharing tips and by sending motivational text messages. Earlier research showed that text messages can be very effective for providing reminders or feedback for achieving a behavioral change [28].

The second aim of the study was to evaluate whether the online program supported patients after treatment in maintaining their changed behaviors and preventing relapses in avoidance and pain-control behaviors. The results of the pilot evaluation showed that, most of all, the motivational messages and the exercises were evaluated as very useful and pleasant. Results from the pilot evaluation showed that 9 patients out of 14 (64%) used the online program after treatment, 5 (5/9, 56%) of whom indicated that the program supported them after treatment. In general, the adherence rates and the evaluation of the program were disappointing. Although it was recommended to use the program daily, the 9 participants out of 14 (64%) that actually used the program only used it once a week or 1 to 5 times in total. Furthermore, 5 out of 14 (36%) participants never used the program. Of these 5 nonusers, 3 (60%) of them indicated that they will use the program in the near future, but we do not know whether these participants really intended to use the program or whether they were providing socially desirable answers. Our results corroborate with earlier results on low adherence rates in online programs for chronic pain patients [6,16]. Daily use of the online relapse-prevention program is essential to assess high-risk situations and to monitor the degree of values-based behavior. Earlier research has shown that there is a relationship between adherence and effect in psychological online interventions [17]. Also in this study, the participants

with positive evaluations did use the program more often than the participants with negative evaluations.

A possible explanation for the low adherence in this study is that most patients had an initial preference with using the program during treatment, particularly with regard to registration of values and values-based actions. For further use, we recommend integrating the program into the multidisciplinary pain treatment program and to train health care professionals to inform and motivate patients to use the program during treatment. Furthermore, the program is not yet successful in providing sufficient social support. The 4 participants out of 14 (29%) who stated that the program was not helpful in the period after treatment indicated that they missed interacting with other participants. Therefore, it is advisable to implement a function where users can react to the shared tips.

Limitations

Limitations of this study included the small samples of users involved in each step and the recruitment of all participants from the same rehabilitation center. Although generalization has to be made with care, the number of participants concurs with the recommended numbers by other studies [29,30]. Another limitation is that we did not involve all stakeholders in each step, such as patients, care providers, managers, and information and communications technology (ICT) developers [21]. However, we worked together closely with the rehabilitation center by informing all the relevant care providers about the results of the steps and by involving the health care professionals in the development process. Finally, in the evaluation of the online program, we did not examine the number of participants that experienced residual problems before and after the relapse-prevention program. For a more effective trial of the online relapse-prevention program, it is important to monitor the effects of the program on relapses.

Future Work and Conclusions

Relapse prevention is still a topic that is hardly examined in the area of chronic pain [4]. This study strengthens the call for further attention for this topic as almost all participants in each step of the development process indicated that they would use a relapse-prevention program after their treatment. Although we provided some insight into the processes of operationalization and summative evaluation of the roadmap, it is relevant to improve a number of conditions in the rehabilitation center for full implementation, such as training health care professionals and integrating the online program into the treatment. An important next step is to evaluate the online relapse-prevention program based on its effectiveness at maintaining the positive effects after multidisciplinary treatment and preventing relapses. Besides measuring relevant health-related measures, it is important to further examine the usage. This will provide information on how to redesign or refine the eHealth technology for achieving higher effects. To conclude, this study provided an overview of how one can design a new, innovative, online relapse-prevention program and revealed valuable insights into the adaptations that have to be made to successfully implement the program in health care.

Acknowledgments

The authors would like to thank the Ministry of Health, Welfare and Sport of the Netherlands for funding this project.

Conflicts of Interest

None declared.

References

1. Smout MF, Hayes L, Atkins PWB, Klausen J, Duguid JE. The empirically supported status of acceptance and commitment therapy: An update. Clin Psychol 2012 Nov 04;16(3):97-109 [FREE Full text] [doi: 10.1111/j.1742-9552.2012.00051.x]
2. Veehof MM, Oskam MJ, Schreurs KM, Bohlmeijer ET. Acceptance-based interventions for the treatment of chronic pain: A systematic review and meta-analysis. Pain 2011 Mar;152(3):533-542 [FREE Full text] [doi: 10.1016/j.pain.2010.11.002] [Medline: 21251756]
3. Williams AC, Eccleston C, Morley S. Psychological therapies for the management of chronic pain (excluding headache) in adults. Cochrane Database Syst Rev 2012;14(11).
4. Morley S. Relapse prevention: Still neglected after all these years. Pain 2008 Feb;134(3):239-240. [doi: 10.1016/j.pain.2007.12.004] [Medline: 18164134]
5. Maruta T, Malinchoc M, Offord KP, Colligan RC. Status of patients with chronic pain 13 years after treatment in a pain management center. Pain 1998 Feb;74(2-3):199-204. [Medline: 9520234]
6. Moessner M, Schiltenwolf M, Neubauer E. Internet-based aftercare for patients with back pain-a pilot study. Telemed J E Health 2012;18(6):413-419. [doi: 10.1089/tmj.2011.0221] [Medline: 22676398]
7. Griffiths F, Lindenmeyer A, Powell J, Lowe P, Thorogood M. Why are health care interventions delivered over the Internet? A systematic review of the published literature. J Med Internet Res 2006;8(2):e10 [FREE Full text] [doi: 10.2196/jmir.8.2.e10] [Medline: 16867965]
8. 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]

9. Bender JL, Radhakrishnan A, Diorio C, Englesakis M, Jadad AR. Can pain be managed through the Internet? A systematic review of randomized controlled trials. Pain 2011 Aug;152(8):1740-1750. [doi: 10.1016/j.pain.2011.02.012] [Medline: 21565446]
10. Macea DD, Gajos K, Daglia Calil YA, Fregni F. The efficacy of Web-based cognitive behavioral interventions for chronic pain: A systematic review and meta-analysis. J Pain 2010 Oct;11(10):917-929. [doi: 10.1016/j.jpain.2010.06.005] [Medline: 20650691]
11. Buhrman M, Fredriksson A, Edström G, Shafiei D, Tärnqvist C, Ljótsson B, et al. Guided Internet-delivered cognitive behavioural therapy for chronic pain patients who have residual symptoms after rehabilitation treatment: Randomized controlled trial. Eur J Pain 2013 May;17(5):753-765. [doi: 10.1002/j.1532-2149.2012.00244.x] [Medline: 23139021]
12. Hayes SC, Luoma JB, Bond FW, Masuda A, Lillis J. Acceptance and commitment therapy: Model, processes and outcomes. Behav Res Ther 2006 Jan;44(1):1-25. [doi: 10.1016/j.brat.2005.06.006] [Medline: 16300724]
13. McCracken LM. Learning to live with the pain: Acceptance of pain predicts adjustment in persons with chronic pain. Pain 1998 Jan;74(1):21-27. [Medline: 9514556]
14. Hayes SC, Strosahl KD, Wilson KG. Acceptance and Commitment Therapy: The Process and Practice of Mindful Change. 2nd edition. New York, NY: The Guilford Press; 2011.
15. Marlatt GA, Donovan DG, editors. Relapse Prevention: Maintenance Strategies in the Treatment of Addictive Behaviors. 2nd edition. New York, NY: The Guilford Press; May 2005.
16. Kristjánsdóttir OB, Fors EA, Eide E, Finset A, Stensrud TL, van Dulmen S, et al. A smartphone-based intervention with diaries and therapist-feedback to reduce catastrophizing and increase functioning in women with chronic widespread pain: Randomized controlled trial. J Med Internet Res 2013;15(1):e5 [FREE Full text] [doi: 10.2196/jmir.2249] [Medline: 23291270]
17. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
18. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
19. Kujala S. User involvement: A review of the benefits and challenges. Behav Inf Technol 2003 Jan;22(1):1-16. [doi: 10.1080/01449290301782]
20. Gould JD, Lewis C. Designing for usability: Key principles and what designers think. Commun ACM 2014 May 14;28(3):300-311. [doi: 10.1145/3166.3170]
21. 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]
22. Kelders SM, Pots WT, Oskam MJ, Bohlmeijer ET, van Gemert-Pijnen JE. Development of a Web-based intervention for the indicated prevention of depression. BMC Med Inform Decis Mak 2013;13(26) [FREE Full text] [doi: 10.1186/1472-6947-13-26] [Medline: 23425322]
23. Trompetter HR, Bohlmeijer ET, Veehof MM, Schreurs KMG. Internet-based guided self-help intervention for chronic pain based on Acceptance and Commitment Therapy: A randomized controlled trial. J Behav Med 2014 Jun 13. [doi: 10.1007/s10865-014-9579-0] [Medline: 24923259]
24. Jaspers MW. A comparison of usability methods for testing interactive health technologies: Methodological aspects and empirical evidence. Int J Med Inform 2009 May;78(5):340-353. [doi: 10.1016/j.ijmedinf.2008.10.002] [Medline: 19046928]
25. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
26. DeLone WH, McLean ER. The DeLone and McLean model of information systems success: A ten-year update. Journal of management information systems 2003;19(4):9-30.
27. van Uden-Kraan CF, Drossaert CH, Taal E, Seydel ER, van de Laar MA. Participation in online patient support groups endorses patients' empowerment. Patient Educ Couns 2009 Jan;74(1):61-69. [doi: 10.1016/j.pec.2008.07.044] [Medline: 18778909]
28. Webb TL, Joseph J, Yardley L, Michie S. Using the Internet to promote health behavior change: A systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
29. Kinzie MB, Cohn WF, Julian MF, Knaus WA. A user-centered model for Web site design: Needs assessment, user interface design, and rapid prototyping. J Am Med Inform Assoc 2002;9(4):320-330 [FREE Full text] [Medline: 12087113]
30. Nielsen J. Estimating the number of subjects needed for a thinking aloud test. Int J Hum Comput Stud 1994 Sep;41(3):385-397. [doi: 10.1006/ijhc.1994.1065]

Abbreviations

ACT: acceptance and commitment therapy

CBT: cognitive behavioral therapy
ICT: information and communications technology
SMS: short message service

Edited by G Eysenbach; submitted 05.02.14; peer-reviewed by C Burton, O Kristjansdottir, M Buhrman; comments to author 17.07.14; revised version received 10.09.14; accepted 08.11.14; published 05.01.15.

Please cite as:
Fledderus M, Schreurs KMG, Bohlmeijer ET, Vollenbroek-Hutten MMR
Development and Pilot Evaluation of an Online Relapse-Prevention Program Based on Acceptance and Commitment Therapy for Chronic Pain Patients
JMIR Human Factors 2015;2(1):e1
URL: http://humanfactors.jmir.org/2015/1/e1/
doi:10.2196/humanfactors.3302
PMID:27025550

Original Paper

Usability Testing of an Internet-Based e-Counseling Platform for Adults With Chronic Heart Failure

Ada YM Payne[1,2], PhD; Jelena Surikova[1], BA; Sam Liu[1,2], MSc, RKin; Heather Ross[2,3], MD, MHSc, FRCPC, FACC; Teodora Mechetiuc[4]; Robert P Nolan[1,2], PhD

[1]Cardiac eHealth, Peter Munk Cardiac Centre, University Health Network, Toronto, ON, Canada

[2]Faculty of Medicine, University of Toronto, Toronto, ON, Canada

[3]Ted Rogers Centre of Excellence in Heart Function, Peter Munk Cardiac Centre, University Health Network, Toronto, ON, Canada

[4]Department of Psychology, McGill University, Montreal, QC, Canada

Corresponding Author:
Ada YM Payne, PhD
Cardiac eHealth
Peter Munk Cardiac Centre
University Health Network
585 University Ave, 6N-617 NU
Toronto, ON, M5G 2N2
Canada
Phone: 1 340 4800 ext 6400
Fax: 1 340 3162
Email: ada.payne@uhnresearch.ca

Abstract

Background: Chronic heart failure (CHF) is a major cause of hospitalization and mortality. In order to maintain heart function and quality of life, patients with CHF need to follow recommended self-care guidelines (ie, eating a heart healthy diet, exercising regularly, taking medications as prescribed, monitoring their symptoms, and living a smoke-free life). Yet, adherence to self-care is poor. We have developed an Internet-based e-Counseling platform, Canadian e-Platform to Promote Behavioral Self-Management in Chronic Heart Failure (CHF-CePPORT), that aims to improve self-care adherence and quality of life in people with CHF. Before assessing the efficacy of this e-platform in a multisite, double-blind, randomized controlled trial, we evaluated the usability of the prototype website.

Objective: The objective of the study was to assess the usability of the CHF-CePPORT e-Counseling platform in terms of navigation, content, and layout.

Methods: CHF patients were purposively sampled from the Heart Function Clinic at the Peter Munk Cardiac Center, University Health Network, to participate in this study. We asked the consented participants to perform specific tasks on the website. These tasks included watching self-help videos and reviewing content as directed. Their interactions with the website were captured using the "think aloud" protocol. After completing the tasks, research personnel conducted a semi-structured interview with each participant to assess their experience with the website. Content analysis of the transcripts from the "think aloud" sessions and the interviews was conducted to identify themes related to navigation, content, and layout of the website. Descriptive statistics were used to summarize the satisfaction data.

Results: A total of 7 men and women (ages 39-77) participated in 2 iterative rounds of testing. Overall, all participants were very satisfied with the content and layout of the website. They reported that the content was helpful to their management of CHF and that it reflected their experiences in coping with CHF. The layout was professional and friendly. The use of videos made the learning process entertaining. However, they experienced many navigation errors in the first round of testing. For example, some participants were not sure how to navigate across a series of Web pages. Based on the experiences that were reported in the first round, we made several changes to the navigation structure. This included using large navigation buttons to direct users to each section and providing tutorial videos to familiarize users with our website. We assessed whether these changes improved user navigation in the second round of testing. The major finding is that participants made fewer navigation errors and they did not identify any new problems.

Conclusions: We found evidence to support the usability of our CHF-CePPORT e-Counseling platform. Our findings highlight the importance of a clear and easy-to-follow navigation structure on user experience.

(JMIR Human Factors 2015;2(1):e7) doi:10.2196/humanfactors.4125

KEYWORDS

chronic heart failure; self-care behaviors; e-counseling; usability assessment

Introduction

Chronic Heart Failure

Chronic heart failure (CHF) is a progressive clinical syndrome in which the heart is unable to pump oxygenated blood sufficiently to meet the metabolic demands of the body during exercise or at rest [1]. It is a major cause of hospitalization and mortality, and it is increasing in prevalence [2]. Prognosis is poor among patients who survive an index admission for CHF, with the 30-day hospital readmission rate at 35% [3]. The 5-year mortality rate is 45% for women and 60% for men [4]. Since there is no cure for CHF, quality of life is a clinically meaningful outcome for these patients [5].

Quality of life is a subjective, multidimensional construct that includes physical, social, and mental well-being [6,7]. CHF symptoms such as shortness of breath and fatigue [8] decrease functional capacity of patients, thereby impeding the pursuit of their life goals and reducing their quality of life [8,9]. Self-care behaviors (ie, recommended guidelines for heart healthy diet, regular exercise, medications, fluid and sodium intake restriction, symptom monitoring, and smoke-free living) are commonly prescribed to manage CHF and to improve quality of life [8]. However, long-term adherence to self-care has been low [10].

In an effort to improve quality of life, our research program has focused on developing telehealth and Internet-based counseling interventions (COHRT, I-START, and REACH) [11-17] that educate and motivate cardiac patients to adopt and maintain self-care behaviors. Our current trial, Canadian e-Platform to Promote Behavioral Self-Management in Chronic Heart Failure (CHF-CePPORT) [18], evaluates the efficacy of an e-Counseling platform in promoting self-care and quality of life among patients with CHF.

Overview of Canadian e-Platform to Promote Behavioral Self-Management in Chronic Heart Failure e-Counseling Platform

This e-Counseling platform has been previously described [18]. In brief, it is an Internet-based preventive e-counseling protocol for patients with CHF. After logging onto the e-platform, users can access content that promotes: (1) explicit validation of the stage of "readiness" for behavior change, (2) active participation in the e-platform via self-guided navigation, (3) commitment to change by using "change talk" [19] to resolve ambivalence and reinforce motivation [20], (4) self-monitoring of behaviors identified by users as a priority for change, and (5) development of cognitive-behavioral skills to build and strengthen efficacy [21] as users embark on their behavior change. The counseling and educational content is reinforced through the use of multimedia, which we will describe below.

The content is organized into 28 e-sessions, delivered over a 12-month period. The e-platform proactively sends out 28 scheduled emails to inform users that new content is available. Each e-session consists of the four core features: (1) self-help video, which connects users with our CHF experts or other patients, and that reflects and validates the experiences of CHF patients; (2) educational content, which provides self-help information that supports users to best manage their condition; (3) interactive e-tools, which help to develop and strengthen self-care behaviors; and (4) e-trackers, which enable self-monitoring of behavior change. Users are self-guided to complete each e-session. A progress graph informs users of the proportion of the e-platform they have completed. They are encouraged to revisit any e-sessions they have previously accessed.

The Need for Usability Assessment

Having a high quality, user-centered program would help maximize engagement and adherence to the e-Counseling platform [22]. To ensure that the e-platform is user-friendly, we conduced a usability assessment, which asked a sample of CHF patients to use the e-platform to perform predetermined tasks under a controlled condition while their experience was documented [23].

Other Internet-based self-management programs have conducted usability studies to help refine their prototype. For example, Stinson et al assessed the usability of an Internet-based, self-management program for adolescents with arthritis and their parents [24]. Voncken-Brewster et al assessed the usability of a Web-based behavioral self-management program for people with chronic obstructive pulmonary disease [25]. The above studies demonstrate that users can help to identify issues related to website design and functionalities that the program developers may have overlooked. Such findings can inform the refinement of the program and maximize its usability when it is fully deployed [22].

Objective

This study examined the usability (navigation, content, layout, and satisfaction) of the CHF-CePPORT e-Counseling platform.

Methods

Study Design

We employed an iterative design [23,26], involving successive rounds of participants in this study. Feedback from the first round (hereafter known as Round 1) of participants was used to inform adjustments to the e-Counseling platform and then

the revised version was tested in the second round (hereafter known as Round 2).

Participants

We used inclusion criteria similar to the CHF-CePPORT trial to recruit participants: (1) male and female patients ≥ 18 years of age; (2) diagnosed with systolic CHF; and (3) fluent in English. We sampled individuals with varying degrees of experience with computers and the Internet to ensure that the e-Counseling platform was easy to use for both novice and advanced computer and Internet users.

Research assistants identified eligible participants from the Heart Function Clinic at the Peter Munk Cardiac Centre, University Health Network. We approached these individuals in person to introduce the study, solicit their consent to participate, and schedule an in-person study session with those who consented.

Procedure

This study received ethics approval by the Research Ethics Board at the University Health Network. Each study visit was divided into two sections: (1) goal-oriented tasks and (2) feedback interview. The overall study visit took up to 1.5 hours.

Before each goal-oriented task, a research assistant read the instructions aloud. The participants were asked to "think-aloud" [27] as they completed each task. This protocol allowed us to directly capture the ongoing thought processes of the participants while using the program, as well as any difficulties they experienced [28]. Participants practiced the think-aloud protocol as they completed a set of sample tasks: they retrieved a nonpersonal standardized email from a sample email account and clicked on a hyperlink in that email. The link redirected them to the log-in page of the e-Counseling platform that we evaluated in this study. None of the participants reported any problems with the think-aloud protocol.

We asked participants to complete two goal-oriented tasks [23]. These tasks allowed us to identify any navigation problems in specific areas of the e-Counseling platform. The first task involved logging onto the website, watching an introductory video, and reading about the e-platform environment. This task assessed the ability of the users to navigate the e-Counseling platform (eg, using hyperlinks to move between pages) and to use an embedded video player. The second task involved completing an e-session that offered self-help tips and tools on active living with CHF. This sample e-session contained 4 core features: (1) a self-help video of exercise experts discussing self-help tips on living an active lifestyle and exercising regularly, (2) educational information that elaborates on the self-help tips mentioned in the above video, (3) an e-tool that helped users to set-up an exercise plan, and (4) a self-monitoring e-tracker of daily step counts. This task tested the ability of users to follow the self-guided session plan and to use the interactive e-tools and e-trackers without assistance.

The research assistants did not offer any help during the goal-oriented tasks, unless explicitly requested by individual participants [27]. This helped to minimize any disruptions to the spontaneous thoughts generated by the participants during task execution. We audiotaped all think-aloud sessions using a digital audio-recorder and then transcribed them verbatim. In addition, another research assistant acted as an observer and documented any problems that the participants may have encountered during the tasks [27]. Both think-aloud transcriptions and field notes were used during data analysis.

After completing the goal-oriented tasks, the research assistant interviewed each participant using a semi-structured interview guide, adopted from Stinson et al [24]. The interview questions focused on the website experience of the participants in three areas: navigation, content, and layout. The interview was also audiotaped on a digital audio-recorder and then transcribed verbatim for analysis. Finally, we asked participants to complete a demographics form and a satisfaction questionnaire. The items on the satisfaction questionnaire were based on the usability characteristics described by Nielsen [29].

Data Analysis

The same protocol was used to analyze data from Rounds 1 and 2. We transcribed the audiotapes and verified the accuracy of the transcripts using a 2-person team: a research assistant transcribed the audiotape verbatim and another independently compared the transcript with the audiotape to check for accuracy.

We conducted a content analysis of the transcripts to identify issues in the three key areas of the e-Counseling platform: navigation, content, and layout. The authors (AP and JS) independently identified and categorized interview excerpts that described: (1) "successful navigation", an incident when a participant was able to follow the website directions correctly or did not experience any problems using the website, for example, able to use a hyperlink to go to the next page, (2) "navigation errors", an incident when a participant was unable to follow the directions provided on the website to complete an e-session or a feature of an e-session. For example, a user began an e-session on the wrong page, unsure of where to go after reviewing a page, or being unable to use a program feature even with instructions provided, and (3) positive and negative comments on various aspects of the website. Coding discrepancies were discussed and resolved between the two coders. Once the coding process was completed, the frequency count of each category was tallied. Pseudonyms were used when reporting any interview excerpts. Means (SD), and percentages were calculated for demographic and satisfaction data.

Results

Participant Description

There were seven individuals who participated in this study, with 4 participants in Round 1, and 3 in Round 2. The sample sizes were determined by data saturation [30]; that is, after the first 4 sessions, we did not identify additional unique issues raised by these participants, and therefore, we concluded Round 1. Similarly, after 3 sessions in Round 2, the participants did not experience the issues reported by those in Round 1, nor experienced any new issues. Thus, we concluded the study with 7 participants.

There were five men and 2 women who completed this study. The age of participants ranged from 39 to 77 years (mean 57, SD 14). There were 5/7 (71%) of them that were married and 6/7 (86%) that completed postsecondary education. There were 5/7 participants (71%) who self-identified as Caucasians, 1/7 (14%) as African American, and 1/7 (14%) as Chinese. Only 1/7 (14%) individual was currently employed.

All participants used computers and the Internet at home. There were 5/7 (71%) of them who were considered “intense users”, spending more than 5 hours per week on the Internet [31]. The individual who was working full-time also used computers and the Internet at work. On a scale from 1 (not at all comfortable) to 5 (very comfortable), Round 1 participants reported a mean comfort level of 4.5 (SD 0.6) with computers and a mean of 5.0 (SD 0.0) with the Internet, while Round 2 participants reported a mean comfort level of 3.3 (SD 1.2) with computers and a mean of 4.0 (SD 1.7) with the Internet. Table 1 provides more detailed descriptions of the 2 samples.

Table 1. Demographics and familiarity with computer and the Internet.

Demographics		Round 1 (n=4)		Round 2 (n=3)	
		n (%)	mean (SD)	n (%)	mean (SD)
Age (years)					
	30-45	1 (25)		1 (33)	
	46-60	1 (25)		1 (33)	
	>60	2 (50)		1 (33)	
Gender					
	Male	3 (75)		2 (67)	
	Female	1 (25)		1 (33)	
Marital status					
	Married/common-law	3 (75)		2 (67)	
	Single	1 (25)		1 (33)	
Highest education level					
	High school	1 (25)		0 (0)	
	College	1 (25)		2 (67)	
	Undergraduate degree	2 (50)		1 (33)	
Current employment status					
	Full-time	0 (0)		1 (33)	
	Disability/leave of absence	1 (25)		0 (0)	
	Unemployed	0 (0)		1 (33)	
	Retired	3 (75)		1 (33)	
Ethnic background					
	Caucasian	3 (75)		2 (67)	
	African-American	1 (25)		0 (0)	
	Chinese	0 (0)		1 (33)	
Computer and Internet usage					
	Do you use the computer at home?				
	Yes	4 (100)		3 (100)	
	No	0 (0)		0 (0)	
	Do you use Internet at home?				
	Yes	4 (100)		3 (100)	
	No	0 (0)		0 (0)	
	Do you use the computer at work?				
	Yes	0 (0)		1 (33)	
	Not applicable	4 (100)		2 (67)	
	How many hours do you spend on the computer each week?				
	≤ 5	0 (0)		2 (67)	
	> 5	4 (100)		1 (33)	
	How many hours do you spend on the Internet each week?				
	≤ 5	0 (0)		2 (67)	
	> 5	4 (100)		1 (33)	

Demographics		Round 1 (n=4)		Round 2 (n=3)	
		n (%)	mean (SD)	n (%)	mean (SD)
	How comfortable are you with using the computer?[a]		4.5 (0.6)		3.3 (1.2)
	How comfortable are you with using the Internet?[a]		5.0 (0.0)		4.0 (1.7)

[a] 1=not at all comfortable, 5=very comfortable

Usability Findings

The study findings are organized into the following themes: overall satisfaction and general comments, navigation, content, and layout.

Overall Satisfaction and General Comments

Table 2 summarizes the results of the satisfaction survey. Participants in both rounds were satisfied with the website, with all items having a mean score of 4 or above on a 5-point rating scale (1=disagree very much; 5=agree very much). There were minor differences on the M ratings between Rounds 1 and 2 participants on their mean item ratings.

Table 2. User satisfaction assessment.

	Round 1 mean (SD)[a]	Round 2 mean (SD)[a]
I learned how to use this website quickly and easily	4.5 (0.6)	4.0 (0.0)
I can find the information I am looking for on this website with no problems	4.8 (0.5)	4.0 (0.0)
I can go through all the materials in an e-session with no problems	4.3 (0.1)	4.3 (0.6)
I am confident that I can remember how to get around this website on my own every time I log on	4.8 (0.5)	4.0 (1.0)
If I get lost on this website, I am confident that I can find my way again	4.8 (0.5)	4.7 (0.6)
I am satisfied with this website	4.8 (0.5)	5.0 (0.0)
I would use this website regularly to help me better manage my heart condition	4.8 (0.5)	4.7 (0.6)

[a] Rating scale, 1=disagree very much; 5=agree very much

Overall Feedback From Participants

Overall, all participants in both rounds of testing were very positive about the e-Counseling platform (see Table 3). They acknowledged the value of having self-care information accessible around the clock not only for new CHF patients, but also long-term patients. The time required to complete an e-session was deemed reasonable. They explored the website freely and believed that this website would be accessible to novice computer and Internet users. There were 2/7 (29%) participants that commented that the website felt like a personalized program. There were 3/7 (43%) participants that expressed that the website gave them hope that they can also live a heart healthy life, as one participant [P3, Round 2] said, "...this [website made] living healthy real to me. And achievable to me."

Table 3. Content analysis, frequency counts of comments under each theme, and navigation issues.

Analyses	Round 1				Round 2			
	# C/I[a]	# of UCs[b]	# of P[c]	mean # C/I per P[d]	# C/I[a]	# of UCs[b]	# of P[c]	mean # C/I per P[d]
Content								
Positive comments	62	27	4	15.5	39	15	3	13.0
Negative comments	9	4	2	4.5	4	3	2	2.0
Navigation								
Positive comments	16	16	4	4.0	30	4	3	10.0
Negative comments	18	18	3	6.0	1	1	1	1.0
Layout								
Positive comments	9	8	4	2.3	9	5	3	3.0
Negative comments	6	3	2	3.0	1	1	1	1.0
User navigation								
Correct navigation	70	14	4	17.5	67	10	3	22.3
Navigation error	24	12	4	6.0	4	3	2	2.0

[a] number of comments or incidents

[b] number of unique comments

[c] number of participants reported

[d] mean number of comments or incidents per participants

User Navigation

We identified a mean of 18 incidents of successful navigation per participant in Round 1 (see Table 3). These participants successfully logged in and out of the website, started the e-session, scrolled down the pages to review the content, navigated to the subsequent pages, and watched embedded videos. However, they also experienced a mean of 6 navigation errors per participant. There were 2/4 (50%) participants that wanted to make the video play on full-screen, but did not know how to do so. There were 3/4 participants (75%) that were uncertain about where to go next during the e-session. All 4 (100%) participants in Round 1 had some difficulties using the unfamiliar interactive features; for example, they did not know how to enter data into the e-tracker. Nevertheless, all the participants understood the purpose behind the interactive features and were willing to try them. They provided some suggestions to improve ease of use, especially for individuals who are less computer-savvy. These suggestions included providing more explicit directions on how to navigate the website.

We modified the website based on these suggestions after Round 1, so that we could evaluate if the changes improved user navigation in Round 2. These changes included using large navigation buttons, instead of text hyperlinks, to take users to the appropriate content and to display clear notices to let users know when they completed an e-session (Figure 1 shows this). In addition, we added several tutorial videos that taught the participants how to use e-tools and e-trackers in order to improve users' familiarity with these features.

Once these changes were made to the website, we observed improvements in navigation in the subsequent round. In Round 2, we identified a mean of 22 incidents of successful navigations and 2 navigation errors per person (see Table 3). There were 2/3 participants (67%) that experienced minor problems: one needed a reminder from the research assistant to scroll down the page for more content and the other did not enter a goal for step count in the e-tracker as instructed during the goal-oriented task. Ultimately, all participants in both rounds agreed that unlike other websites in which you can navigate freely, there is a learning curve to the self-guided structure of the CHF-CePPORT e-Counseling platform. However, the participants reported that they would have no problem navigating the website if they had a chance to use it once or twice at home. A participant [P2, Round 1] said, "...it will take a little time to get mastered and get on top of it, but I suppose any new website is like that and [this website is] a lot easier and more straightforward than any I've seen." Since these last two minor issues did not significantly impede the successful completion of an e-session, we did not make further navigation-related modifications.

When we examined the comments related to navigation, we identified a mean of 4 positive comments and 6 negative comments per participant in Round 1 (see Table 3). In Round 2, participants made a mean of 10 positive comments per person and only 1 individual made a negative comment about navigation. Table 4 provides sample comments.

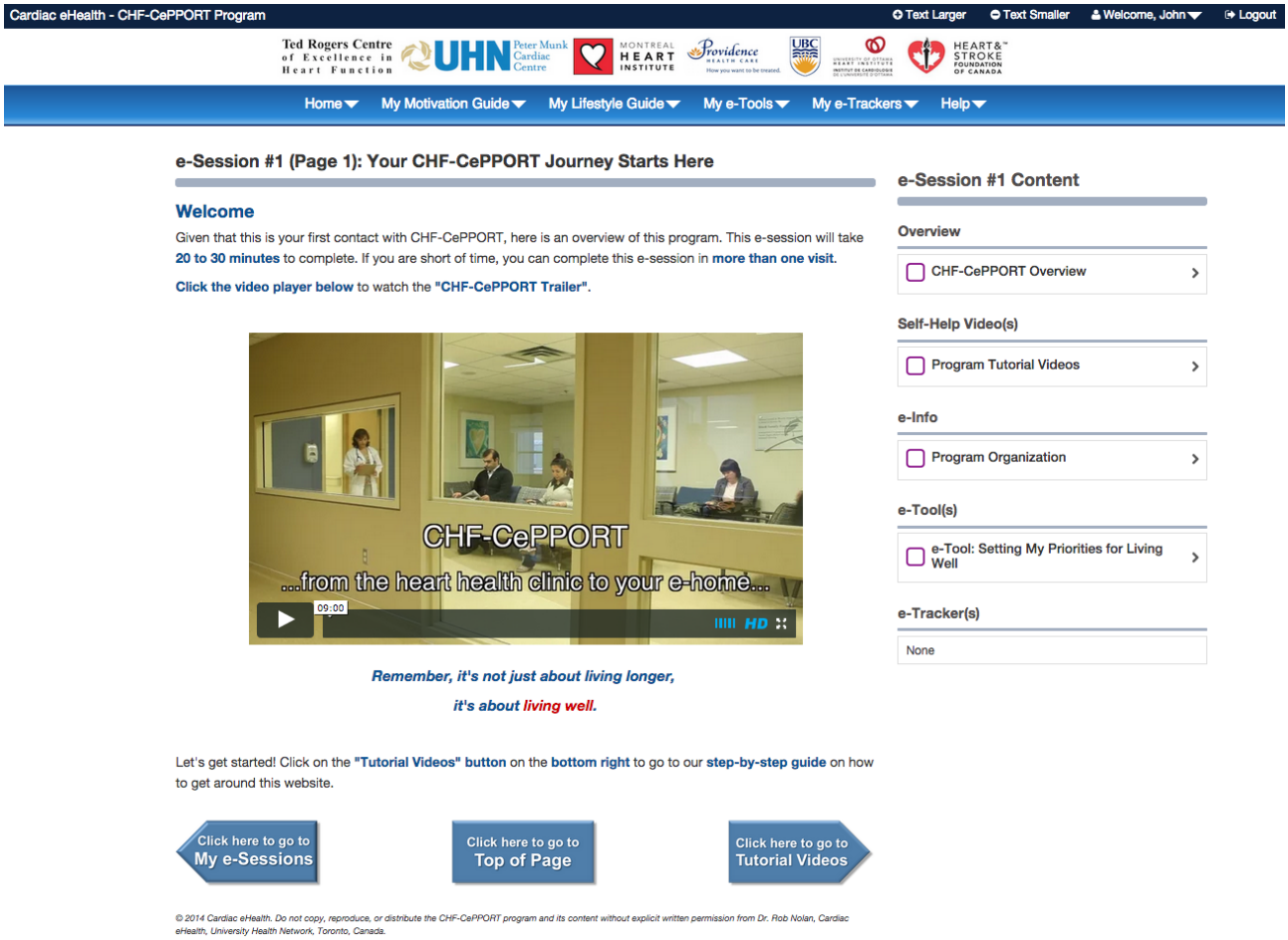
Table 4. Sample comments for each of the themes.

Theme and its definition	Example of a positive comment	Example of a negative comment
General comment:[a] Comments made about the overall website	• "...living with heart problem, I'm not by myself. There is something that could really help me."[P3, Round 2] • "I know that you're doing this for a vast amount of people, but it really feels like this has been catered to me personally..."[P4, Round 1]	---[b]
Navigation: The ability for participants to independently move around the website, review the content, and use the e-tools and e-trackers as designed	• "...for the first time, it's figuring out, most of the time it told you that...the bottom right hand side of the screen to go to the next step and push the [button] and it took you to the next step...to me, that was important to know where the button was. Because sometimes they don't tell ya and you're looking, where is it?"[P1, Round 2]	• "...hopefully I navigated the right way because it's still not, from my perspective, completely intuitive as to once you come on to [the website], where you need to go."[P1, Round]
Content: The material offered by the website, which can include self-help videos, didactic information, interactive e-tool, and self-monitoring e-trackers	• "What I read most of was the exercise part and that was very helpful, very straight forward and the step counting and everything was very informative and it was very clear and quite complete and you have charts for schedules, which is a help. I found that very direct and straightforward."[P2, Round 1] • "...this is a guide, this for a self-help situation...what is being put here is to reinforce what I've already been told, what we've already been told [about self-care behaviors]."[P2, Round 2]	• "...in this interview scenario, [the experts] are not facing you...if [they] were actually talking to the participant instead of each other that might be a better way of engaging someone?"[P3, Round 1]
Layout: The visual appearance of the website, including color, font size, images	• "...it is a very attractive site. Yes, the pictures and it's not just words and it's well set up and arranged in charts so I think it's very good...visually..."[P2, Round 1]	• "...it might be a good feature maybe to have a font size where you can make it a little larger..."[P1, Round 2]

[a] Individual participants were identified by a subject number and the testing round in which they participated.

[b] All of the comments were deemed to be positive by both coders.

Figure 1. A screenshot of Canadian e-Platform to Promote Behavioral Self-Management in Chronic Heart Failure (CHF-CePPORT) e-Counseling platform after the second round of usability testing. Original image.



Content

We identified a mean of 16 positive and 5 negative comments per participant in Round 1. In Round 2, we identified a mean of 13 positive and 2 negative comments per person in Round 2 (see Table 3).

Participants were very positive about the content of the e-Counseling platform, as indicated by the greater number of positive versus negative comments from both rounds of participants. All participants commented that the self-help materials were helpful, straightforward, approachable, and practical for cardiac patients and their families. The content covered a wide range of topics and reinforced the importance of adherence to self-care behaviors. Furthermore, the material addressed experiences faced by heart failure patients. A patient [P1, Round 2] said, "...[the content is] realistic to what you actually go through as a heart patient, the struggles you have going through..."

Participants also commented that videos and interactive tools helped to make the information easier to understand and the learning process more entertaining. There were 5/7 (71%) participants that thought the videos are of good quality. The experts in the videos offered useful information in an empathetic manner, which participants appreciated. There were 1/7 (14%) participant that commented that the daily step count e-tracker was a helpful tool to monitor self-care behavior change, while 3/7 (43%) participants appreciated having the progress graph to keep track of their progression through the CHF-CePPORT e-platform. To improve ease of use, 1/7 (14%) participant recommended the use of multiple-choice and check-box response options for some of the interactive tools. We plan to incorporate this suggestion in the next generation of the e-platform.

Layout

We identified a mean of 2 positive and 3 negative comments per person about the layout in Round 1. We identified a mean of 3 positive comments per person regarding the website layout and only 1 negative comment in Round 2 (see Table 3).

Comments on the layout of the website were mostly positive in both rounds of testing. The participants remarked that the website had a clear and professional look. Its appearance was warm and friendly without being distracting to participants. A participant [P3, Round 2] remarked, "...the meat is the real thing. The real knowledge [on] how to take care of myself...instead of something flashing, or some colorful thing [to] distract me." There were 2/7 (29%) participants that wondered if the font size was large enough for others with visual impairment. To address this, we subsequently added instructions, accessible at the top of every page, on how to adjust the font size through the Web browser.

Discussion

Principal Findings

The goals of this usability study were to gather feedback on the CHF-CePPORT e-Counseling platform and identify any navigation issues that may impede its usage. Although the findings highlighted some issues that our team did not anticipate, especially related to navigation, we also received many positive comments about the layout and content of the e-platform. This information was critical to the implementation of the e-Counseling platform in our multicenter, double-blind, randomized controlled trial [18]. By identifying and addressing usability challenges, we can ensure that these issues are unlikely to confound our trial findings.

Overall, the participants were very satisfied with the e-Counseling platform. Participants reported that it was easy to review the content and find the information for which they were looking. Most of our participants commented that the videos on our e-platform were of high quality and made them pay attention to the content. The learning process was entertaining and not time-consuming. This feedback was encouraging and supported our use of "edutainment" [32], for example, films that simulate real-life situations for patient education. This technique has been shown to improve accessibility and understanding of complex medical information (eg, medical tests and treatment options) among people of various levels of health literacy [33-35].

Our goal was to create an e-platform that is user-friendly for an older population because a significant portion of CHF patients are 65 years or older [36]. This segment of the population tends to be less computer literate than a younger age group, despite their growing engagement with the Internet [31]. Yet, they are also avid consumers of health information on the Web. Our Web design features incorporated recommendations made by the National Institute on Aging and the National Library of Medicine to ensure accessibility and ease of use [37]. These included using larger font size, white spaces around text, and a simple color scheme to improve readability. In this study, the participants were very positive about our layout and design. They commented that it was professional looking, while conveying warmth and friendliness. These participants did not have trouble reading the text, though a couple of them suggested offering a way to adjust the font size. Based on this feedback, we believe that the current e-platform design is suitable for older individuals with CHF.

A study goal was to identify and address any navigation issues. This was critical because such issues can impede users from accessing the clinical content, thus minimizing the effectiveness of the e-Counseling platform [38]. There were fewer navigation errors and negative comments made during Round 2. As a result, we believe that the changes we have made to the website after Round 1 have improved the accessibility of our website. This finding is even more encouraging because this improvement was observed from the participants in Round 2, who felt less comfortable with computers and the Internet than those in Round 1.

Limitations

A few limitations should be noted. First, we relied on observations made by research personnel to document the interactions between participants and the e-platform. Using other methods such as video-capture of mouse clicks and screen display would provide more objective data on such interactions. However, the involvement of multiple independent data coders enhanced the rigor of our data analysis and interpretation process. Second, we tested the usability of the e-platform using a sample e-session. There are a total of 28 e-sessions in the CHF-CePPORT e-platform. We chose one of the e-sessions for testing to minimize respondent burden. Although this sample e-session included all the core features of the e-platform, there may be other usability issues in the remaining 27 e-sessions that we have not identified. Last, our sample of 7 participants may seem insufficient in assessing the usability of the e-platform. However, 80%-95% of usability problems can be identified using 5-9 individuals [39]. Moreover, we did not uncover additional issues from Round 2 participants. Thus, we are confident that the majority of usability problems on the e-Counseling platform have been identified and addressed.

Conclusions

In this study, we found evidence to support the usability of the CHF-CePPORT e-Counseling platform. In addition to the content and layout, navigation proved to be a critical component for the design of our website. Internet-based self-management programs are becoming more common as a way to compliment medical therapies to manage complex diseases such as CHF. The CHF-CePPORT e-platform is consistent with our priority to design and implement an easy-to-follow navigational structure to facilitate user access.

Acknowledgments

The CHF-CePPORT trial is funded by an operating grant (#MOP-287717) from the Canadian Institutes of Health Research, as well as a research grant from the Peter Munk Cardiac Centre Innovation Fund (#5790-5090-0602). Thank you to the participants for their contribution to our study.

Conflicts of Interest

The authors of this paper are also the developers of the Internet-based e-counseling intervention being tested in the CHF-CePPORT trial.

References

1. Liu P, Arnold M, Belenkie I, Howlett J, Huckell V, Ignazewski A, et al. The 2001 Canadian Cardiovascular Society consensus guideline update for the management and prevention of heart failure. Can J Cardiol 2001 Dec;17 Suppl E:5E-25E. [Medline: 11773943]
2. Bui AL, Horwich TB, Fonarow GC. Epidemiology and risk profile of heart failure. Nat Rev Cardiol 2011 Jan;8(1):30-41 [FREE Full text] [doi: 10.1038/nrcardio.2010.165] [Medline: 21060326]
3. Dharmarajan K, Hsieh AF, Lin Z, Bueno H, Ross JS, Horwitz LI, et al. Diagnoses and timing of 30-day readmissions after hospitalization for heart failure, acute myocardial infarction, or pneumonia. JAMA 2013 Jan 23;309(4):355-363 [FREE Full text] [doi: 10.1001/jama.2012.216476] [Medline: 23340637]
4. Malcom J, Arnold O, Howlett JG, Ducharme A, Ezekowitz JA, Gardner MJ, et al. Canadian Cardiovascular Society Consensus Conference guidelines on heart failure--2008 update: Best practices for the transition of care of heart failure patients, and the recognition, investigation and treatment of cardiomyopathies. Can J Cardiol 2008 Jan;24(1):21-40 [FREE Full text] [Medline: 18209766]
5. Johansson P, Dahlström U, Broström A. Factors and interventions influencing health-related quality of life in patients with heart failure: A review of the literature. Eur J Cardiovasc Nurs 2006;5(1):5-15. [doi: 10.1016/j.ejcnurse.2005.04.011]
6. Geneva: World Health Organization. 1997. Measuring quality of life: The World Health Organization quality of life instruments (The WHOQOL-100 and the WHOQOL-BREF) URL: http://www.who.int/mental_health/media/68.pdf [accessed 2015-04-23] [WebCite Cache ID 6Y0OtKGUb]
7. Maciver J, Ross HJ. Quality of life and left ventricular assist device support. Circulation 2012 Aug 14;126(7):866-874 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.111.040279] [Medline: 22891167]
8. Heo S, Lennie TA, Okoli C, Moser DK. Quality of life in patients with heart failure: Ask the patients. Heart Lung 2009;38(2):100-108 [FREE Full text] [doi: 10.1016/j.hrtlng.2008.04.002] [Medline: 19254628]
9. Jeon YH, Kraus SG, Jowsey T, Glasgow NJ. The experience of living with chronic heart failure: A narrative review of qualitative studies. BMC Health Serv Res 2010;10:77 [FREE Full text] [doi: 10.1186/1472-6963-10-77] [Medline: 20331904]
10. Evangelista LS, Shinnick MA. What do we know about adherence and self-care? J Cardiovasc Nurs 2008;23(3):250-257 [FREE Full text] [doi: 10.1097/01.JCN.0000317428.98844.4d] [Medline: 18437067]
11. Angus JE, Rukholm E, Michel I, Larocque S, Seto L, Lapum J, et al. Context and cardiovascular risk modification in two regions of Ontario, Canada: A photo elicitation study. Int J Environ Res Public Health 2009 Sep;6(9):2481-2499 [FREE Full text] [doi: 10.3390/ijerph6092481] [Medline: 19826558]
12. Grace SL, Barry-Bianchi S, Stewart DE, Rukholm E, Nolan RP. Physical activity behavior, motivational readiness and self-efficacy among Ontarians with cardiovascular disease and diabetes. J Behav Med 2007 Feb;30(1):21-29. [doi: 10.1007/s10865-006-9080-5] [Medline: 17109217]
13. Nolan RP, Upshur RE, Lynn H, Crichton T, Rukholm E, Stewart DE, et al. Therapeutic benefit of preventive telehealth counseling in the Community Outreach Heart Health and Risk Reduction Trial. Am J Cardiol 2011 Mar 1;107(5):690-696. [doi: 10.1016/j.amjcard.2010.10.050] [Medline: 21215382]
14. Durrani S, Irvine J, Nolan RP. Psychosocial determinants of health behaviour change in an e-counseling intervention for hypertension. Int J Hypertens 2012;2012:191789 [FREE Full text] [doi: 10.1155/2012/191789] [Medline: 22229083]
15. Nolan RP, Liu S, Shoemaker JK, Hachinski V, Lynn H, Mikulis DJ, et al. Therapeutic benefit of internet-based lifestyle counselling for hypertension. Can J Cardiol 2012 May;28(3):390-396. [doi: 10.1016/j.cjca.2012.02.012] [Medline: 22498181]
16. Walker RL, Gee ME, Bancej C, Nolan RP, Kaczorowski J, Joffres M, et al. Health behaviour advice from health professionals to Canadian adults with hypertension: Results from a national survey. Can J Cardiol 2011;27(4):446-454. [doi: 10.1016/j.cjca.2011.03.013] [Medline: 21684718]
17. Nolan RP, Liu S, Feldman R, Dawes M, Barr S, Lynn H, et al. Reducing risk with e-based support for adherence to lifestyle change in hypertension (REACH): Protocol for a multicentred randomised controlled trial. BMJ Open 2013;3(8):e003547 [FREE Full text] [doi: 10.1136/bmjopen-2013-003547] [Medline: 23965936]
18. Nolan RP, Payne AY, Ross H, White M, D'Antono B, Chan S, et al. An Internet-based counseling intervention with email reminders that promotes self-care in adults with chronic heart failure: Randomized controlled trial protocol. JMIR Res Protoc 2014;3(1):e5 [FREE Full text] [doi: 10.2196/resprot.2957] [Medline: 24480783]
19. Miller WR, Rose GS. Toward a theory of motivational interviewing. Am Psychol 2009 Sep;64(6):527-537 [FREE Full text] [doi: 10.1037/a0016830] [Medline: 19739882]
20. Miller WR, Rollnick S, Conforti K. Motivational interviewing: Preparing people for change. USA: The Guilford Press; 2002.
21. 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]
22. Ahern DK. Challenges and opportunities of eHealth research. Am J Prev Med 2007 May;32(5 Suppl):S75-S82. [doi: 10.1016/j.amepre.2007.01.016] [Medline: 17466822]
23. Wichansky AM. Usability testing in 2000 and beyond. Ergonomics 2000 Jul;43(7):998-1006. [doi: 10.1080/001401300409170] [Medline: 10929833]

24. Stinson J, McGrath P, Hodnett E, Feldman B, Duffy C, Huber A, et al. Usability testing of an online self-management program for adolescents with juvenile idiopathic arthritis. J Med Internet Res 2010;12(3):e30 [FREE Full text] [doi: 10.2196/jmir.1349] [Medline: 20675293]
25. Voncken-Brewster V, Moser A, van der Weijden T, Nagykaldi Z, de VH, Tange H. Usability evaluation of an online, tailored self-management intervention for chronic obstructive pulmonary disease patients incorporating behavior change techniques. JMIR Res Protoc 2013;2(1):e3 [FREE Full text] [doi: 10.2196/resprot.2246] [Medline: 23612363]
26. Currie LM. Evaluation frameworks for nursing informatics. Int J Med Inform 2005 Dec;74(11-12):908-916. [doi: 10.1016/j.ijmedinf.2005.07.007] [Medline: 16099711]
27. Someren M, Barnard YF. The think aloud method: A practical guide to modelling cognitive processes. London: Academic Press; 1994.
28. Jaspers MW, Steen T, van den Bos Cor C, Geenen M. The think aloud method: A guide to user interface design. Int J Med Inform 2004 Nov;73(11-12):781-795. [doi: 10.1016/j.ijmedinf.2004.08.003] [Medline: 15491929]
29. Nielsen J. Iterative user-interface design. Computer 1993 Nov;26(11):32-41. [doi: 10.1109/2.241424]
30. Morse JM. The Significance of saturation. Qual Health Res 1995 May 01;5(2):147-149. [doi: 10.1177/104973239500500201]
31. Veenhof B, Timusk P. Online activities of Canadian boomers and seniors: Statistics Canada. Ottawa: Statistics Canada; 2014. URL: http://www.statcan.gc.ca/pub/11-008-x/2009002/article/10910-eng.htm [accessed 2014-12-13] [WebCite Cache ID 6UmNEJkti]
32. Nariman HN. Soap operas for social change: Toward a methodology for entertainment-education television. Westport, Conn: Praeger; 1993.
33. Jibaja-Weiss ML, Volk RJ. Utilizing computerized entertainment education in the development of decision aids for lower literate and naïve computer users. J Health Commun 2007;12(7):681-697. [doi: 10.1080/10810730701624356] [Medline: 17934944]
34. Jibaja-Weiss ML, Volk RJ, Granchi TS, Neff NE, Robinson EK, Spann SJ, et al. Entertainment education for breast cancer surgery decisions: A randomized trial among patients with low health literacy. Patient Educ Couns 2011 Jul;84(1):41-48. [doi: 10.1016/j.pec.2010.06.009] [Medline: 20609546]
35. Volk RJ, Jibaja-Weiss ML, Hawley ST, Kneuper S, Spann SJ, Miles BJ, et al. Entertainment education for prostate cancer screening: A randomized trial among primary care patients with low health literacy. Patient Educ Couns 2008 Dec;73(3):482-489 [FREE Full text] [doi: 10.1016/j.pec.2008.07.033] [Medline: 18760888]
36. Jugdutt BI. Aging and heart failure: Changing demographics and implications for therapy in the elderly. Heart Fail Rev 2010 Sep;15(5):401-405. [doi: 10.1007/s10741-010-9164-8] [Medline: 20364319]
37. National Institute on Aging/the National Library of Medicine. Making your website senior friendly: A checklist. Washington, DC: US Government Print Office; 2001. URL: http://www.nia.nih.gov/sites/default/files/making_your_website_senior_friendly.pdf [accessed 2015-04-23] [WebCite Cache ID 6Y0PEWhLL]
38. Farkas DK, Farkas JB. Guidelines for designing web navigation. Technical Communication. 2000. URL: https://hcde.uw.edu/files/people/docs/farkasfarkas-guidelinesfornavigation.pdf [accessed 2015-04-23] [WebCite Cache ID 6Y0PUPxDB]
39. Virzi RA. Refining the test phase of usability evaluation: How many subjects is enough? Human Factors: The Journal of the Human Factors and Ergonomics Society 1992;34:457-468. [doi: 10.1177/001872089203400407]

Abbreviations

CHF: chronic heart failure
CHF-CePPORT: Canadian e-Platform to Promote Behavioral Self-Management in Chronic Heart Failure

Edited by G Eysenbach; submitted 17.12.14; peer-reviewed by M Jibaja Weiss; comments to author 03.03.15; revised version received 06.03.15; accepted 22.03.15; published 08.05.15.

Please cite as:
Payne AYM, Surikova J, Liu S, Ross H, Mechetiuc T, Nolan RP
Usability Testing of an Internet-Based e-Counseling Platform for Adults With Chronic Heart Failure
JMIR Human Factors 2015;2(1):e7
URL: http://humanfactors.jmir.org/2015/1/e7/
doi:10.2196/humanfactors.4125
PMID:27026267

Original Paper

Nurses' Perceptions and Practices Toward Clinical Alarms in a Transplant Cardiac Intensive Care Unit: Exploring Key Issues Leading to Alarm Fatigue

Azizeh Khaled Sowan[1], RN, PhD; Albert Fajardo Tarriela[2], RN; Tiffany Michelle Gomez[3], RN, MN; Charles Calhoun Reed[3], RN, PhD; Kami Marie Rapp[4], RN

[1]University of Texas Health Science Center at San Antonio, School of Nursing, Department of Health Restoration and Care Systems Management, San Antonio, TX, United States

[2]University Health System, Transplant Cardiac ICU, San Antonio, TX, United States

[3]University Health System, Center for Nursing Excellence, San Antonio, TX, United States

[4]University Health System, Nursing Administration, San Antonio, TX, United States

Corresponding Author:
Azizeh Khaled Sowan, RN, PhD
University of Texas Health Science Center at San Antonio
School of Nursing, Department of Health Restoration and Care Systems Management
School of Nursing
7703 Floyd Curl Dr. - MC 7975
San Antonio, TX, 78229-3900
United States
Phone: 1 210 567 5799
Fax: 1 210 567 1719
Email: sowan@uthscsa.edu

Abstract

Background: Intensive care units (ICUs) are complex work environments where false alarms occur more frequently than on non-critical care units. The Joint Commission National Patient Safety Goal .06.01.01 targeted improving the safety of clinical alarm systems and required health care facilities to establish alarm systems safety as a hospital priority by July 2014. An important initial step toward this requirement is identifying ICU nurses' perceptions and common clinical practices toward clinical alarms, where little information is available.

Objective: Our aim was to determine perceptions and practices of transplant/cardiac ICU (TCICU) nurses toward clinical alarms and benchmark the results against the 2011 Healthcare Technology Foundation's (HTF) Clinical Alarms Committee Survey.

Methods: A quality improvement project was conducted on a 20-bed TCICU with 39 full- and part-time nurses. Nurses were surveyed about their perceptions and attitudes toward and practices on clinical alarms using an adapted HTF clinical alarms survey. Results were compared to the 2011 HTF data. Correlations among variables were examined.

Results: All TCICU nurses provided usable responses (N=39, 100%). Almost all nurses (95%-98%) believed that false alarms are frequent, disrupt care, and reduce trust in alarm systems, causing nurses to inappropriately disable them. Unlike the 2011 HTF clinical alarms survey results, a significantly higher percentage of our TCICU nurses believed that existing devices are complex, questioned the ability and adequacy of the new monitoring systems to solve alarm management issues, pointed to the lack of prompt response to alarms, and indicated the lack of clinical policy on alarm management (P<.01). Major themes in the narrative data focused on nurses' frustration related to the excessive number of alarms and poor usability of the cardiac monitors. A lack of standardized approaches exists in changing patients' electrodes and individualizing parameters. Around 60% of nurses indicated they received insufficient training on bedside and central cardiac monitors. A correlation also showed the need for training on cardiac monitors, specifically for older nurses (P=.01).

Conclusions: False and non-actionable alarms continue to desensitize TCICU nurses, perhaps resulting in missing fatal alarms. Nurses' attitudes and practices related to clinical alarms are key elements for designing contextually sensitive quality initiatives to fight alarm fatigue. Alarm management in ICUs is a multidimensional complex process involving usability of monitoring devices, and unit, clinicians, training, and policy-related factors. This indicates the need for a multi-method approach to decrease alarm fatigue and improve alarm systems safety.

(JMIR Human Factors 2015;2(1):e3) doi:10.2196/humanfactors.4196

KEYWORDS

clinical alarms; alarm fatigue; critical care; physiologic monitors; nursing; survey

Introduction

Clinical alarms are the top hazard listed in the 2014 Emergency Care Research Institute's (ECRI) "Top Ten Health Technology Hazards" report [1]. Consensus exists on research about the low specificity and excessive number of false alarms (86%-99.5%) produced by physiological monitors [2-4]. This results in clinicians ignoring or disabling alarms, a phenomenon known as alarm fatigue, and raises a question about the clinical value of the currently used physiological alarm systems.

Fatal incidents related to clinical alarms are well documented [5,6]. As a result, the Joint Commission (JC) National Patient Safety Goal (NPSG .06.01.01) targeted improving the safety of clinical alarm systems, requiring health care facilities to establish alarm systems safety as a hospital priority by July 2014 [7].

A 2005-2006 national survey of more than 1300 health care professionals and other hospital personnel (such as monitor technicians and clinical engineers) by the Healthcare Technology Foundation (HTF) Clinical Alarms Committee showed that nuisance alarms are frequent (81%), disrupt care (77%), and reduce trust in alarms, causing clinicians to inappropriately disable them (78%) [8]. Newer-monitoring systems did not solve alarms problems (69%) [8]. The HTF clinical alarms survey is the most comprehensive survey available to date, in comparison to other surveys measuring perceptions and attitudes toward clinical alarms [9,10]. The HTF clinical alarms survey was developed by a group of multidisciplinary experts in biomedical engineers, safety, and instrumentation and was supported for administration by different safety and regulatory agencies such as the Association for the Advancement of Medical Instrumentation, Food and Drug Administration/MEDSUN, ECRI, and others [8]. Surprisingly, a 2011 administration of the same survey to a larger group of clinicians (N=4278) revealed very similar results [11]. It is worth noting that the combined samples of 2005-2006 and 2011 HTF clinical alarms surveys were from more than ten different hospital departments, including the intensive care units (ICUs).

Monitoring the physiological condition of critically ill patients is a complex task where the use of multiple monitoring devices per patient is the norm and an essential component in the treatment process. Nurses are the key professionals responding to alarms and managing the multiple monitoring devices [5]. The high rate of false alarms constantly reported in ICUs [12-14] compared to non-critical care units [15] resulted in nurses responding to an average of 150-400 alarms per patient per day in ICUs [16]. Therefore, alarm safety is a clear priority in these units. This also suggests that ICU nurses may have different perceptions toward clinical alarms than nurses in other clinical areas. Thus, nurses' attitudes toward clinical alarms and their perceptions of factors that may threaten alarm recognition and response are essential in guiding research projects and quality initiatives for alarm management in ICUs.

None of the available studies on perceptions and attitudes to alarm management have yet to benchmark their results with the HTF clinical alarms national data using the complete version of the survey. Little information is available about ICU nurses' attitudes and common practices related to clinical alarms [9,10,17], and none is available specifically about nurses' perceptions and attitudes in transplant/cardiac ICU (TCICU), the target setting of this project. Examining nurses' attitudes and practices toward clinical alarms using a comprehensive survey such as the HTF clinical alarms survey is essential to understand the complexity of the ICU work environment and contributing factors that threaten the safety of alarm recognition and appropriate management. Additionally, in the project setting, state-of-the-art new physiological monitoring devices are used, demanding an evaluation of their capabilities to reduce alarm fatigue and improve alarm safety.

Nurses on a 20-bed TCICU identified an excessive number of clinical alarms, specifically from the cardiac monitors, as a safety hazard that caused work disruption. An interprofessional alarm management taskforce consisting of nurses, physicians, and biomedical engineers was assembled to attain phase 1 A. of the JC NPSG.06.01.01, to establish alarm systems safety as a hospital priority. The initial project goal was to standardize alarm management in all ICUs. This phase of the project had two objectives: (1) to determine TCICU nurses' perceptions and attitudes toward clinical alarms signal from all physiological monitors, as well as current practices and educational needs for alarm management using the cardiac monitors, and (2) to benchmark the results with the 2011 HTF clinical alarms survey data. Correlations between attitudes, nurse characteristics, and other factors such as training on monitoring devices were also examined for further insight into the current problem.

Methods

Design, Sample, and Setting

Approval to conduct this quality improvement project was obtained from the hospital's Institutional Review Board, and implied consent was obtained from the participating nurses. This project was conducted on a 20-bed TCICU located in a 684-bed university teaching Magnet hospital in the Southwestern United States. The unit has 39 full- and part-time nurses with a nurse patient ratio of 1:2. The unit is equipped with modern patient monitoring devices (eg, cardiac monitors, pulmonary artery catheter monitoring, pulse oximeter) and with intensive care equipment for life support (eg, ventilator, infusion pump). The ICU is an "E" shape with patients' rooms to the sides and an unstaffed central monitor station. At the time of this project, the unit had no policy for clinical alarm management.

In April 2014, the unit witnessed two major changes: the implementation of new cardiac monitors (Philips IntelliVue MX800) and Wi-Fi (CISCO) phones for communication. When deploying any physiological monitoring devices, nurses usually

receive a group-based presentation with hands-on training on the device's appropriate use by the device company representative. Device manuals are also available in the unit for nurses to review. Newly hired nurses are trained on device use during their orientation program by their preceptors, unit educators, or the company representatives. Usually, no other structured periodic training on managing physiological monitoring devices is offered. However, the nursing unit educators do provide individualized help for device management if needed. This project began 2.5 months after implementing the new cardiac monitors.

Instrument and Procedure

We adapted the 2011 HTF clinical alarms survey after obtaining approval from the developers for its use to understand TCICU nurses' attitudes and practices related to clinical alarms. Four expert ICU nurses reviewed the survey after adaptation for face validity and appropriateness to use in the TCICU. The adapted survey included three sections: (1) demographics, (2) perception about clinical alarms signaled from all monitoring devices, and (3) potential issues that interfere with alarm recognition. Section 2 in the HTF clinical alarms survey had 20 statements rated using a 5-point Likert-type scale of agreement followed by a free-text area to provide details on statements. Our changes to the HTF clinical alarms survey items involved Section 2 and consisted of (1) deleting the statement "The integration of clinical alarms into the Joint Commission patient safety measures have reduced patient adverse events" because it is not applicable to our setting yet, (2) adding three statements to capture other alarm issues specific to the TCICU related to the types of alarms, alarm specificity, and the unit layout, and (3) replacing "institution" and "floor/area of the hospital" to "unit" in some statements, to reflect the context of measurement. Section 3 has 9 issues to order in rank from 1 (most important) to 9 (least important).

Since the HTF clinical alarms survey is designed to measure clinicians' attitudes toward alarms signaled from all physiological monitoring devices and because the unit deployed new cardiac monitors, three additional questions were also asked to understand nurses' practices toward clinical alarms specific to the cardiac monitors. These were related to the (1) frequency of individualizing alarms' parameters, (2) frequency of changing electrodes, and (3) adequacy of the training received on using the cardiac monitors.

The survey was designed using SurveyMonkey and placed on a hospital website. In coordination with the nursing director of the TCICU, we sent individual recruitment emails, each with a unique ID, to all 39 TCICU nurses with a link to the adapted survey. ID numbers were used for follow-up on responses. Two email reminders were sent to non-respondents by the first author (AS), who is not employed at the project setting, to enhance the response rate.

Data Analysis

Descriptive statistics were used to describe respondents' characteristics (age, clinical, and computer experience) and to summarize questionnaire responses. A *Z* test for difference in proportions for independent samples was used to examine the difference between the percentages of HTF clinical alarms survey respondents and TCICU nurses in this project. Mean ranks were used for the ranking section of the survey. Content analysis was used to categorize the narrative data into themes. Bivariate correlations between demographic information and other survey statements and questions related to training and practices were calculated using a chi-square test. A level of significance of .05 was used for all statistical tests.

Results

Overview

A total of 39 completed responses (100% response rate) were obtained with usable data. The majority of nurses were females (25/39, 64%), about 40 years old (28/39, 72%), and full-time staff (33/39, 85%). The percentages of nurses who reported having "1-3" and ">5" years of overall nursing experience were equal (16/39, 41%, in each category). However, the majority of nurses were within 1-3 years of TCICU experience (28/39, 72%). The mean score of the reported computer skills was 2.4 (SD 0.68) out of a 4-point Likert-type scale.

General Statements About Clinical Alarms

Table 1 presents the percentages of TCICU nurses and HTF clinical alarms study respondents who agreed/strongly agreed on each of the 22-item statements about clinical alarms. A major assumption of *Z* test is that "n*p and n(1-p) must both be equal to or greater than 5", where n is the sample size and p is the proportion. This assumption was not met when a very high percentage of our participants agreed/strongly agreed with several item statements; therefore, *Z* test for the difference in proportions was not calculated for the first five items.

Similar to the majority of the HTF study sample, almost all of our TCICU nurses agreed or strongly agreed on the first five statements regarding the frequency of nuisance alarms and the need for distinctive alarm sounds and visual displays (Table 1). The majority of the respondents from the two studies were also supportive of the use of smart alarms, hiring dedicated central monitor alarm management staff, and integrating the alarms into wireless devices (Items 6, 7, 10, 11). Almost two thirds of our TCICU nurses indicated that the unit layout interferes with alarm recognition and management (Item 8), and only half agreed that lethal alarms are responded to promptly (Item 15).

In contrast to the HTF study results, a significantly higher percentage of our nurses pointed to confusion in locating an alarming device (Item 9), believed that existing devices are complex for setting alarms parameters (Item 12), questioned (disagreed with) the ability and adequacy of the monitoring systems to alert staff of changes in a patient's condition (Item 14), doubted the sensitivity of the clinical staff to alarms (Item 17), and did not think that the monitoring devices provided distinct outputs (Item 19). Additionally, the majority of our nurses indicated a lack of requirements to document the individualization of patient parameters (Item 20) and the absence of clinical policies on alarm management (Item 21). Almost all nurses believed that the new monitoring systems have not solved most of the previous problems they experienced with clinical alarms (Item 22).

Table 1. Percentages of TCICU nurses who agreed or strongly agreed on clinical alarm survey statements compared with respondents of the 2011 HTF survey data.

#	Statement	TCICU n[a] (%)	HTF 2011 n[b] (%)	*P*
1	Nuisance alarms disrupt patient care	38 (98)	4125 (71)	NA[c]
2	Nuisance alarms reduce trust in alarms and cause caregivers to inappropriately turn alarms off at times other than setup or procedural events	38 (98)	4133 (78)	NA[c]
3	Alarm sounds and/or visual displays of the current monitoring systems and devices should clearly differentiate the priority of alarm	37 (95)	4137 (91)	NA[c]
4	Alarm sounds and/or visual displays should be distinct based on the parameter or source (eg, device)	37 (95)	4130 (91)	NA[c]
5	Nuisance alarms occur frequently	37 (95)	4124 (77)	NA[c]
6	Smart alarms (eg, where multiple parameters, rate of change of parameters, and signal quality are automatically assessed in their entirety) would be effective to use for improving clinical response to important patient alarms	31 (80)	3783 (78)	.7
7	Smart alarms (eg, where multiple parameters, rate of change of parameters, and signal quality are automatically assessed in their entirety) would be effective to use for reducing false alarms	30 (78)	3791 (78)	.9
8[d]	Unit layout does interfere with alarm recognition and management	28 (73)	NA	NA
9	When a number of devices are used with a patient, it can be confusing to determine which device is in an alarm condition	28 (73)	3916 (51)	<.01[e]
10	Central alarm management staff responsible for receiving alarm messages and alerting appropriate staff is helpful	24 (59)	3890 (53)	.4
11	Alarm integration and communication systems via pagers, cell phones, and other wireless devices are useful for improving alarms management and response	23 (56)	3786 (56)	.9
12	Properly setting alarm parameters and alerts is overly complex in existing devices	22 (56)	4009 (21)	<.001[e]
13	Environmental background noise has interfered with alarm recognition	21 (54)	3919 (42)	.1
14[f]	The alarms used on my unit are adequate to alert staff of potential or actual changes in a patient's condition	20 (51)	3978 (72)	<.001[e]
15[d]	When a lethal alarm sounds, it is clearly and quickly recognized and immediate action is taken to address the alarm	19 (49)	NA	NA
16[d]	Nearly all alarms are actionable (requiring the nurse to respond and take an action)	19 (49)	NA	NA
17	Clinical staff is sensitive to alarms and responds quickly	13 (34)	3935 (66)	<.001[e]
18	There have been frequent instances where alarms could not be heard and were missed	12 (32)	3999 (29)	.6
19[f]	The medical devices used on my unit all have distinct outputs (ie, sounds, repetition rates, visual displays) that allow users to identify the source of the alarm	12 (32)	3927 (70)	<.001[e]
20[f]	There is a requirement in my unit to document that the alarms are set and are appropriate for each patient	12 (29)	3784 (71)	<.001[e]
21[f]	Clinical policies and procedures regarding alarm management are effectively used in my unit	8 (20)	3772 (55)	<.001[e]
22	Newer monitoring systems (eg, <3 years old) have solved most of the previous problems we experienced with clinical alarms	1 (2)	3988 (29)	<.001[e]

[a]This "n" reflects only the participants who agreed/strongly agreed on each statement and not the total sample size. The total sample size was 39.

[b]This "n" is the number of respondents who answered each statement and is not limited to those who agreed/strongly agreed on each statement, and was used to calculate *Z* test. These numbers are unpublished data and were obtained from the HTF. The total sample size of the 2011 HTF survey is 4278.

[c]NA= Not applicable. No *Z* scores were calculated for difference between the two studies on these statements because "n*p and n(1-p)" were less than 5.

[d]These are the new statements that we added to our survey and were not available in the HTF survey. Therefore, no *Z* score was calculated.

[e]Significant at $P<.05$.

[f]These are the statements where the "floor/area of the hospital" or "institution" in the HTF clinical alarms survey were replaced with "unit".

Narrative Data

A total of 22 nurses provided narrative comments about clinical alarms and issues threatening timely recognition and response. Categories, themes, and examples of comments are listed in Table 2. All comments were negative reflecting serious issues related to safety; poor usability of the cardiac monitors; a lack of support to the use of evidence-based solutions for alarm management, such as watchers for the central monitors and connecting alarms of the monitoring devices to the communication devices (eg, CISCO phones) [5,18]; and unit-related factors, such as a lack of policy to manage alarms, unit layout interferes with alarm response, and the need for further training on the cardiac monitors.

Table 2. Categories, themes, and comments of the TCICU nurses' narrative data (N=22).

Categories and themes	Examples of comments
Category 1: Frequent false alarms and patient safety	
Theme 1: False alarms are very frequent and very distracting (12 nurses)	"too much alarms that distract care and patient sleep"
	"they signal for no reason even in an empty patient room"
	"the continuous "bing" of the central monitor gives me a huge headache"
Theme 2: There is a tendency by nurses to ignore clinical alarms (5 nurses)	"the nuisance of the new cardiac monitors is so overwhelming you tend to ignore"
	"I have watched multiple nurses at the nursing desk listen to alarms sounding and not respond, very worrisome"
Category 2: Poor usability of the medical devices	
Theme 3: Alarms' sounds and visual displays are not distinct based on the priority of the alarm, parameter, or the device (9 nurses)	"lethal alarms are not distinguishable than other alarms"
	"alarms' sounds and visual displays sound and look alike for different vitals"
Theme 4: The new cardiac monitors are very complex and not user friendly (4 nurses)	"newer cardiac monitors made it worst, they are just fancier"
	"cardiac monitors are too difficult to navigate, and takes away time to care for patient which is more important than figuring the monitor to function, they are FOREVER alarming"
	"I am unable to correct false alarms easily"
	"alarms will sound for false Vtachs with no way to silence or relearn"
	"cardiac monitor can't recognize the waveform of SPO2, adjustment on wave height is necessary"
Theme 5: The lowest volume of the alarms is still very loud and distracts patient sleep (2 nurses)	"alarms are very loud within the room, even turning the volume down to the lowest level is still loud- keeps patients awake at night"
	"the new cardiac monitors have the same volume alarm for even the most trivial alarms that it sets a cry wolf mentality and could pose a dangerous situation in which an actual true alarm could be disregarded"
Category 3: Lack of support to the use of evidence-based solutions for alarm management	
Theme 6: A central monitor watcher will not solve the problem (3 nurses)	"having a watcher might be unsafe, will relax the monitoring eyes/ears of a nurse as a another person is equally monitoring"
	"it will add to alarm fatigue, it would be easier for me to just go in the room and fix the problem than have someone constantly calling me"
Theme 7: Unreliable technology to integrate with alarms (3 nurses)	"CISCO phones and pagers sometimes don't alert or receive any alarms even for emergencies, there are delays on them and they loose the signals in the elevators"
Category 4: Unit-related factors to alarm management	
Theme 8: Absence of alarm management and documentation policy (3 nurses)	"we need to reinforce that alarm parameters need to be changed specific to the patient"
	"there is no place in the medical record to document that alarms are individualized based on patient condition"
Theme 9: Unit layout may hinder response to alarms (2 nurses)	"although alarms are loud within the patient room, the E-shape unit makes the unit too large and resulted in alarms being unheard"
	"even within the same hallway a fatal alarm can be missed"
	"with the big unit, we cannot see all patients in the central monitor unless adjustment is done"
Theme 10: Further training on monitoring devices is required (1 nurse)	"there is not enough time to train staff on the central monitor alarm"

Ranking of Issues that Affect Alarm Recognition

Table 3 presents ranking of the issues that may affect alarm recognition and response by TCICU nurses and HTF study respondents. The top four critical issues identified by TCICU nurses endangering alarm recognition and response were similar to the HTF data. However, the rankings of these issues differed according to our nurses who, for example, ranked "difficulty in identifying the source of an alarm" as the first critical issue versus the HTF respondents who ranked this issue as second. Interestingly, and similar to the HTF study, our nurses ranked the lack of training as one of the three least important issues, as well as noise competition from nonclinical alarms.

Table 3. Ranking of TCICU nurses compared to respondents of the 2011 HTF clinical alarms survey on the importance of issues that affect response to alarms (1=most important, 9=least important).

Items	Our ICU data (N=39)		HTF 2011 data (N=4276)	
	Mean[a]	Ranking[b]	Mean[a]	Ranking[b]
Difficulty in identifying the source of an alarm	2.94	1	4.61	2
Difficulty in understanding the priority of an alarm	3.06	2	4.64	3
Difficulty in hearing alarms when they occur	3.93	3	4.70	4
Frequent false alarms, which lead to reduced attention or response to alarms when they occur	4.15	4	4.21	1
Inadequate staff to respond to alarms as they occur	4.23	5	4.87	6
Difficulty in setting alarms properly	4.44	6	5.16	7
Noise competition from nonclinical alarms and pages	4.45	7	5.66	9
Over-reliance on alarms to call attention to patient problems	4.77	8	4.86	5
Lack of training on alarm systems	6.60	9	5.55	8

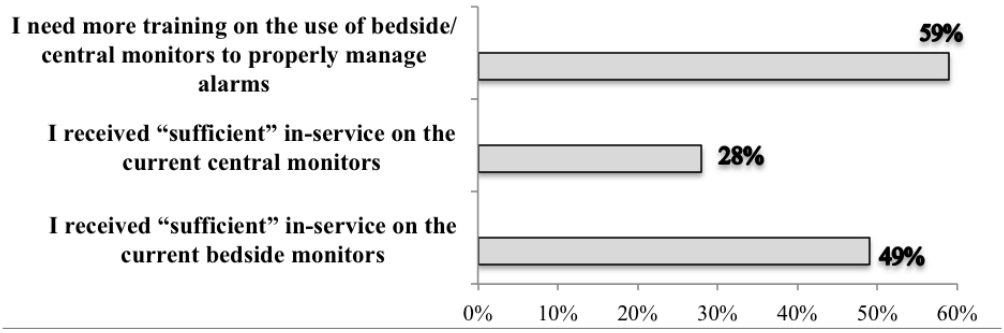
[a]Mean rank of the item.

[b]Ranking of the mean.

Nurses' Practices and Level of Training Related to Cardiac Monitors

The results support the lack of standardized approaches in changing patients' electrodes and individualizing parameters. Only half of the nurses reported changing electrodes every 24 hours (51%, 20/39 nurses). Other nurses reported changing electrodes only when needed (23%, 9/39), every shift (13%, 5/39), or every 48 hours (13%, 5/39). Similarly, more than one third of the nurses indicated not changing monitors' parameters, and only 5% (2/39) change parameters after disconnecting the patient from the monitor and when the setting reverted to defaults. Over half of the nurses indicated the need for more training on the bedside and central cardiac monitors (Figure 1).

Figure 1. Percentages of TCICU nurses who agreed/strongly agreed on the adequacy of the training received on bedside and central cardiac monitors (N=39).



Correlations

Bivariate correlations using chi-square test were examined between age, computer skills, years of ICU experience (as the demographic data), and the perception about the complexity of the monitoring devices (survey Item 12), adequacy of alarms to alert staff (Item 14), frequency of changing patient parameters and electrodes, and the need for further training on cardiac monitors. All variables were recoded as binary. None of the correlations were significant, except for the positive relationship between age and the need for further training on cardiac monitors (P=.01).

Discussion

Overview

A mountain of evidence exists on the need for alarm management [19]. However, the majority of the available studies targeted changing specific parameters or new algorithms and their effect on decreasing the number of false alarms and were

not guided by issues recognized by clinicians as critical for alarm management and response [12-14,20-23]. Identifying key issues leading to alarm desensitization as a safety threat should be a priority in alarm management. This project focused on understanding the context-related attitudes and practices of TCICU nurses toward clinical alarms and benchmarked the results with the 2011 HTF clinical alarms national data. The results showed that alarm fatigue is a critical and urgent issue in our TCICU. In comparison to the HTF study results, the responses of our TCICU nurses highlighted the complexity of alarm management in ICUs.

Principal Findings

This project helped identify key issues leading to alarm desensitization. Our nurses and the HTF study respondents agree that false alarms occur frequently, disrupt care, and reduce trust in alarm systems and response sensitivity, causing clinicians to inappropriately disable them. On the other hand, and unlike the HTF data, the majority of our TCICU nurses challenged the ability and adequacy of the new monitoring systems in solving these alarm management issues. Narrative data attributed this primarily to usability issues with the devices, specifically the cardiac monitors. These were non-trivial issues such as the complexity in navigation to set alarm parameters, the inability of the nurse to turn off some of the false alarms or to adjust alarm volume, failure of the monitors to identify that patients were disconnected from monitors and were alarming in empty rooms, failure of the monitors to display the waveforms of the parameters in their appropriate size, and the look-alike and sound-alike alarms for different parameters with different priorities and from different devices. Most important, these issues affected the timely recognition of lethal alarms, resulting in only 50% of the nurses reporting prompt response to such alarms.

The comments related to poor usability and lack of user-centered devices were all linked to the new cardiac monitors, indicating the need for future research on usability testing even for the newest devices and especially for complex ICU monitoring devices that may jeopardize safety and workflow efficiency. Little information is available about the usability of physiological monitoring devices [24,25]. Unlike previous studies on alarm fatigue [9,10], usability of the monitoring devices was a major reason behind nurses' frustration with alarm systems in this project. These findings are congruent with the fast pace, high-stress level, and complexity of the ICUs, where monitoring devices need to be useful tools to guide clinical decision-making rather than positive contributors to the stress level, workload, workflow inefficiency, and sleep deprivation among patients.

One of the nurses described the cardiac monitor as a "fancy" monitor, suggesting the availability of unused features by clinicians and perhaps the lack of knowledge on the appropriate use and usefulness of some features. For patient safety in ICUs, previous research supported the need to eliminate unnecessary alarms [22] and to understand triggered defaults. In fact, overuse of alarms of the monitoring devices is a practice associated with overdiagnosis; it may do more harm than good. Finding only 5% of our nurses changing alarm parameters after disconnecting the patient from the monitor may indicate that nurses are unaware that the disconnection results in settings reverting to defaults (a critical safety feature).

Results showed that our nurses were supportive of employing a dedicated person for the central monitor and the integration of alarms into the communication devices [5]. However, narrative comments highlighted safety and feasibility concerns. This indicates the need to pilot any initiatives for alarm management to assure their appropriateness.

Another unique finding of this project was that unit layout was a major factor interfering with alarm response and recognition. Instead, unit architectural layout should be a facilitator to timely response and recognition. Nurse input to unit design is imperative in the future. Additionally, alarm policies and requirements to document alarm settings were absent. An alarm management policy could eliminate non-standardized practices related to frequency of changing the electrodes and customizing patients' parameters. The American Association of Critical Care Nurses' evidence-based recommendation of electrode change is "daily and if needed" [26]. Only 50% of our nurses were following this recommendation. Also, more than a third of our nurses reported not customizing alarm parameters to be patient specific. These practices contribute significantly to increasing the number of false alarms in ICUs [5], but examining whether nurses have sufficient knowledge on parameter limits is equally important. While 44% of the TCICU nurses customize the parameters, evidence-based hard stops are needed in these devices, specifically for critical parameters.

The rankings assigned to the importance of issues identified the source of an alarm and understanding its priority as the top two critical issues, reflecting the complexity of the monitoring devices and their current inadequacy. Difficulty in hearing alarms, ranked as the third most important issue, may be attributed to the unit layout as explained by nurses in the narrative comments. The lack of training was listed as the least important factor, but in contrast, 60% of the nurses doubted their abilities to manage cardiac monitors and requested further training. This may be because the training question was limited to the cardiac monitors while survey items concerned all existing devices. Most important, nurses' responses reflect the high frustration level of nurses who think that devices should be designed to be easy to use at a minimum and should help nurses acknowledge the source and priority of the alarms without the nurse spending time figuring out basic operational issues. Furthermore, the need for further training may also reflect deficiencies in the current group-training method, suggesting techniques such as the use of simulation [5], periodic refresher training, and super users. Interestingly, our results also supported a positive correlation between age and the need for training. This indicates that training methods may need to be revised for older nurses or that older nurses might be more resistant to change.

Summary and Future Directions

Our results highlight the complexity of overall alarm management in ICUs and that ICU nurses may have different perceptions toward alarm management than other nurses. Appropriate alarm management depends on a combination of

device usability, training, unit layout, IT infrastructure, and alarm management protocols and documentation capabilities [23,27]. In summary, this complexity suggests that (1) policies should be in place to guide end users of monitoring devices on alarm management, (2) device usability is fundamental for alarm management and emphasis in this area is needed, (3) the traditional group-based, one-round training on complex alarm-equipped monitoring devices is inadequate, (4) a need exists for structured evaluation of quality initiatives to ensure their appropriateness for different work cultures, and (5) focusing on one strategy (eg, changing alarms' algorithms) to decrease false alarms may be insufficient to improve alarm fatigue.

Limitations

The findings of this project can be generalized with caution. We obtained a 100% response rate, indicating a motivated sample, perhaps reflecting the importance of this issue to ICU nurses, the high stress level experienced by our nurses toward clinical alarms, and the need for urgent initiative to manage this problem. However, the sample size is relatively small and the results are limited to a TCICU in one setting with monitors from specific vendors. Other frequently used devices in other ICUs may also contribute negatively or positively to alarm fatigue. Including other ICUs will also increase the sample size. We measured nurses' attitudes 2.5 months after the introduction of the new cardiac monitors because we thought this time period would be sufficient for nurses to adapt to the new devices. Measuring attitudes before or after that time period might reveal other findings as a result of novelty in using the devices (if measured before) or adaptation to the new monitors (if measured after). Last, although we could not use the *Z* test to measure the difference between our nurses and the HTF respondents on the first five statements on the survey; the comparable high percentages of respondents from the two studies on these statements predict the absence of any statistical differences.

Conclusions

Clinical alarm management is in its infancy in many institutes. False and non-actionable alarms continue to desensitize clinicians and may result in missed fatal alarms. A multi-method approach in decreasing alarm fatigue and improving alarm systems safety is needed across devices, training, unit layout, clinicians, and policies. Usability of monitoring devices is essential in alarm management. Clinicians' attitudes and practices related to clinical alarms are key in designing contextually sensitive quality initiatives to fight alarm fatigue. Partnership between clinicians, organizations, researchers, manufacturers, safety, and regulatory organizations is essential to improve alarm management. In the future, a comparison across other ICUs is needed and comprehensive usability studies are essential.

Acknowledgments

This project was funded by the University Health System's Department of Patient Care Services and the Center for Excellence in Patient Care. We thank Dr Nancy Staggers, an adjunct professor at the Department of Biomedical Informatics and the School of Nursing University of Utah, and Dr Cheryl Lehman, a retired professor from the School of Nursing University of Texas Health Science Center San Antonio for reviewing this paper.

Conflicts of Interest

None declared.

References

1. ECRI. ECRI Institute Top 10 Health Technology Hazards Report for 2014 URL: http://www.visimobile.com/wp-content/uploads/2013/12/6C-5-Top-10-Health-Tech-Hazards-2014-%E2%80%93-ECRI.pdf [accessed 2015-01-02] [WebCite Cache ID 6VHMueROe]
2. Imhoff M, Kuhls S. Alarm algorithms in critical care monitoring. Anesth Analg 2006 May;102(5):1525-1537. [doi: 10.1213/01.ane.0000204385.01983.61] [Medline: 16632837]
3. Phillips J, Barnsteiner JH. Clinical alarms: improving efficiency and effectiveness. Crit Care Nurs Q 2005;28(4):317-323. [Medline: 16239820]
4. Atzema C, Schull MJ, Borgundvaag B, Slaughter GRD, Lee CK. ALARMED: adverse events in low-risk patients with chest pain receiving continuous electrocardiographic monitoring in the emergency department. A pilot study. Am J Emerg Med 2006 Jan;24(1):62-67. [doi: 10.1016/j.ajem.2005.05.015] [Medline: 16338512]
5. Association for the Advancement of Medical Instrumentation. Clinical alarms: 2011 summit URL: http://www.aami.org/htsi/alarms/pdfs/2011_Alarms_Summit_publication.pdf [accessed 2015-01-01] [WebCite Cache ID 6VHNSNksF]
6. The Joint Commission. Sentinel event alerts: Medical device alarm safety in hospitals.: April 8, issue 50; 2013. URL: http://www.jointcommission.org/assets/1/18/SEA_50_alarms_4_5_13_FINAL1.PDF [accessed 2015-01-01] [WebCite Cache ID 6VHNuN0Aa]
7. The JC (2014). National Patient Safety Goals Effective January 1, 2014 URL: http://www.jointcommission.org/assets/1/6/HAP_NPSG_Chapter_2014.pdf [accessed 2015-01-01] [WebCite Cache ID 6VHNZo71j]
8. Clinical Alarm Task Force, AACE Healthcare Technology Foundation. Impact of clinical alarms on patient safety: A report from the American College of Clinical Engineering Healthcare Technology Foundation. J Clin Eng 2007;32.

9. Christensen M, Dodds A, Sauer J, Watts N. Alarm setting for the critically ill patient: a descriptive pilot survey of nurses' perceptions of current practice in an Australian Regional Critical Care Unit. Intensive Crit Care Nurs 2014 Aug;30(4):204-210. [doi: 10.1016/j.iccn.2014.02.003] [Medline: 24703797]
10. Graham KC, Cvach M. Monitor alarm fatigue: standardizing use of physiological monitors and decreasing nuisance alarms. Am J Crit Care 2010 Jan;19(1):28-34; quiz 35 [FREE Full text] [doi: 10.4037/ajcc2010651] [Medline: 20045845]
11. Funk M, Clark JT, Bauld TJ, Ott JC, Coss P. Attitudes and practices related to clinical alarms. Am J Crit Care 2014 May;23(3):e9-e18. [doi: 10.4037/ajcc2014315] [Medline: 24786820]
12. Görges M, Markewitz BA, Westenskow DR. Improving alarm performance in the medical intensive care unit using delays and clinical context. Anesth Analg 2009 May;108(5):1546-1552. [doi: 10.1213/ane.0b013e31819bdfbb] [Medline: 19372334]
13. Siebig S, Kuhls S, Imhoff M, Langgartner J, Reng M, Schölmerich J, et al. Collection of annotated data in a clinical validation study for alarm algorithms in intensive care--a methodologic framework. J Crit Care 2010 Mar;25(1):128-135. [doi: 10.1016/j.jcrc.2008.09.001] [Medline: 19327311]
14. Siebig S, Kuhls S, Imhoff M, Gather U, Schölmerich J, Wrede CE. Intensive care unit alarms--how many do we need? Crit Care Med 2010 Feb;38(2):451-456. [doi: 10.1097/CCM.0b013e3181cb0888] [Medline: 20016379]
15. Schmid F, Goepfert MS, Kuhnt D, Eichhorn V, Diedrichs S, Reichenspurner H, et al. The wolf is crying in the operating room: patient monitor and anesthesia workstation alarming patterns during cardiac surgery. Anesth Analg 2011 Jan;112(1):78-83. [doi: 10.1213/ANE.0b013e3181fcc504] [Medline: 20966440]
16. HCPro (2011). From the AHAP blog: Joint Commission and FDA targarm fatigue URL: http://www.hcpro.com/ACC-265320-4634/From-the-AHAP-Blog-Joint-Commission-and-FDA-target-alarm-fatigue.html [accessed 2014-12-31] [WebCite Cache ID 6VHNyhQMO]
17. Cvach MM, Frank RJ, Doyle P, Stevens ZK. Use of pagers with an alarm escalation system to reduce cardiac monitor alarm signals. J Nurs Care Qual 2014;29(1):9-18. [doi: 10.1097/NCQ.0b013e3182a61887] [Medline: 23963169]
18. ECRI (2014). Clinical alarms URL: https://www.ecri.org/components/HRC/Pages/CritCare5.aspx?tab=2 [accessed 2014-12-31] [WebCite Cache ID 6VHd5G8jP]
19. Cvach M. Monitor alarm fatigue: an integrative review. Biomed Instrum Technol 2012;46(4):268-277. [doi: 10.2345/0899-8205-46.4.268] [Medline: 22839984]
20. Burgess LPA, Herdman TH, Berg BW, Feaster WW, Hebsur S. Alarm limit settings for early warning systems to identify at-risk patients. J Adv Nurs 2009 Sep;65(9):1844-1852. [doi: 10.1111/j.1365-2648.2009.05048.x] [Medline: 19694847]
21. Aboukhalil A, Nielsen L, Saeed M, Mark RG, Clifford GD. Reducing false alarm rates for critical arrhythmias using the arterial blood pressure waveform. J Biomed Inform 2008 Jun;41(3):442-451 [FREE Full text] [doi: 10.1016/j.jbi.2008.03.003] [Medline: 18440873]
22. Biot L, Holzapfel L, Becq G, Mélot C, Baconnier P. Do we need a systematic activation of alarm soundings for blood pressure monitoring for the safety of ICU patients? J Crit Care 2003 Dec;18(4):212-216. [Medline: 14691894]
23. Solet JM, Barach PR. Managing alarm fatigue in cardiac care. Progress in Pediatric Cardiology 2012 Jan;33(1):85-90. [doi: 10.1016/j.ppedcard.2011.12.014]
24. Drews FA. Patient monitors in critical care: Lessons for improvement. In: Henriksen K, Battles JB, Keyes MA. editors. Advances in Patient Safety: New Directions and Alternative Approaches (Vol. 3: Performance and Tools). Rockville (MD): Agency for Healthcare Research and Quality (US); 2008.
25. Hyman WA. Human factors: should your medical devices require intensive care? Crit Care Nurs Clin North Am 2010 Jun;22(2):233-241. [doi: 10.1016/j.ccell.2010.03.005] [Medline: 20541072]
26. American Association of Critical Care. 2013. Alarm management - Practice alert URL: http://www.aacn.org/wd/practice/docs/practicealerts/alarm-management-practice-alert.pdf [accessed 2015-01-01] [WebCite Cache ID 6VHOOtZLJ]
27. ECRI Institute. The hazards of alarm overload. Keeping excessive physiologic monitoring alarms from impeding care. Health Devices 2007 Mar;36(3):73-83. [Medline: 17500263]

Abbreviations

ECRI: Emergency Care Research Institute
HTF: Healthcare Technology Foundation
ICU: intensive care unit
JC: Joint Commission
NPSG: National Patient Safety Goal
TCICU: transplant/cardiac ICU

Edited by G Eysenbach; submitted 10.01.15; peer-reviewed by M Cvach, W Hyman; comments to author 05.02.15; revised version received 24.02.15; accepted 25.02.15; published 16.03.15.

Please cite as:
Sowan AK, Tarriela AF, Gomez TM, Reed CC, Rapp KM
Nurses' Perceptions and Practices Toward Clinical Alarms in a Transplant Cardiac Intensive Care Unit: Exploring Key Issues Leading to Alarm Fatigue
JMIR Human Factors 2015;2(1):e3
URL: http://humanfactors.jmir.org/2015/1/e3/
doi:10.2196/humanfactors.4196
PMID:27025940

Original Paper

Nursing Performance and Mobile Phone Use: Are Nurses Aware of Their Performance Decrements?

Deborah McBride[1*], RN, PhD; Sandra A LeVasseur[2*], RN, PhD; Dongmei Li[3*], PhD

[1]Samuel Merritt University, Oakland, CA, United States

[2]Hawaii State Center for Nursing, University of Hawaii at Manoa, Honolulu, HI, United States

[3]Clinical and Translational Science Institute, University of Rochester School of Medicine and Dentistry, Rochester, NY, United States

[*]all authors contributed equally

Corresponding Author:
Deborah McBride, RN, PhD
Samuel Merritt University
3100 Telegraph Ave
Oakland, CA, 94609
United States
Phone: 1 510 848 1721
Fax: 1 650 292 5566
Email: dmcbride@samuelmerritt.edu

Abstract

Background: Prior research has documented the effect of concurrent mobile phone use on medical care. This study examined the extent of hospital registered nurses' awareness of their mobile-phone-associated performance decrements.

Objective: The objective of this study was to compare self-reported performance with reported observed performance of others with respect to mobile phone use by hospital registered nurses.

Methods: In March 2014, a previously validated survey was emailed to the 10,978 members of the Academy of Medical Surgical Nurses. The responses were analyzed using a two-proportion z test (alpha=.05, two-tailed) to examine whether self-reported and observed rates of error were significantly different. All possible demographic and employment confounders which could potentially contribute to self-reported and observed performance errors were tested for significance.

Results: Of the 950 respondents, 825 (8.68%, 825/950) met the inclusion criteria for analysis. The representativeness of the sample relative to the US nursing workforce was assessed using a two-proportion z test. This indicated that sex and location of primary place of employment (urban/rural) were represented appropriately in the study sample. Respondents in the age groups <40 years old were underrepresented, while age groups >55 years old were overrepresented. Whites, American Indians/Alaskan natives, and Native Hawaiian or Pacific Islanders were underrepresented, while Hispanic and multiple/other ethnicities were overrepresented. It was decided to report the unweighted, rather than the weighted survey data, with the recognition that the results, while valuable, may not be generalizable to the entire US registered nursing workforce. A significant difference was found between registered nurses' self-reported and observed rates of errors associated with concurrent mobile phone use in following three categories (1) work performance (z=−26.6142, P<.001, Fisher's exact test), (2) missing important clinical information (z=−13.9882, P=.008, Fisher's exact test), and (3) making a medical error (z=−9.6798, P<.001, Fisher's exact test). Respondents reported that personal mobile phone use by nurses at work was a serious distraction; always (13%, 107/825), often (29.6%, 244/825), sometimes (44.6%, 368/825), rarely (8.7%, 72/825), or never (1.2%, 10/825). On balance, 69.5% (573/825) of respondents believed that nurses' use of personal mobile phones while working had a negative effect on patient care. Since all possible confounders were tested and none were deemed significant, a multivariate analysis was not considered necessary.

Conclusions: Many hospitals are drawing up policies that allow workers to decide how to use their devices at work. This study found that nurses express a disproportionately high confidence in their ability to manage the risk associated with the use of mobile phones and may not be able to accurately assess when it is appropriate to use these devices at work.

(JMIR Human Factors 2015;2(1):e6) doi:10.2196/humanfactors.4070

KEYWORDS

distraction; mobile phone; cellular phone; Internet; nurses; hospital; non-work related mobile phone use

Introduction

Studies examining the effect of mobile phones on work performance have proliferated [1-3], and they have demonstrated that concurrent mobile phone use results in performance decrements, including the impaired ability to focus, difficulty filtering out extraneous information, and the impaired ability to remember important information. In clinical environments, this distraction could cause substandard patient care. Katz-Sidlow et al [4] reported that 19% of residents and 12% of attending physicians acknowledged missing important clinical information because of distractions from mobile phones. In addition, 34% of residents and 20% of attending physicians reported observing another team member miss an important piece of clinical information because they were distracted by their mobile phones during rounds. Smith et al [5] surveyed surgical technicians, of whom 92.7% reported that they had never been distracted by or had their performance negatively affected by their mobile phone, and 98% reported that they had never made an error that could be attributed to their mobile phone use. In contrast, 34.5% reported seeing another surgical technician distracted by their mobile phone during surgery. These results suggest that while many clinicians are aware of the potential dangers of using mobile phones while working, they may not be aware of their own decreased performances. Another study described the underestimation of self-reported medical errors relative to those observed [6]. The researchers identified a self-esteem bias, where respondents have a subconscious tendency to protect themselves from the emotional distress associated with personal deficiencies, resulting in a lack of awareness of medical errors.

The objective of this study was to compare self-reported performance decrements with reports of observed performance decrements related to mobile phone use by registered hospital nurses.

Methods

We used a survey instrument piloted in 2013 [7]. The survey consisted of four parts, with questions about (1) demographics, (2) the use of personal communication devices, (3) opinions about the effects of personal communication devices on the work of registered nurses, and (4) hospital policies concerning personal communication devices (Multimedia Appendix 1). The questions, which were developed based on a literature review and interviews with hospital nurses, asked respondents to rank the types of activities they engage in on a 5-point Likert scale to determine how frequently they participated in each activity. Psychometric testing of the questionnaire included examining internal consistency reliability and test-retest reliability in a sample of 50 registered nurses. A Spearman rho correlation was used to determine the test-retest reliability. There was strong test-retest reliability between the two tests with a mean agreement for the Likert scale responses of 74% (SD 15, range 43-100%). Accounting for responses within 1 SD range on the Likert scale increased the agreement to 96% (SD 7, range 87-100%). The Cronbach coefficient alpha values examining the internal consistency and reliability in three of the domains were high; utilization (.84), impact (.96), and opinion (.85). A lower agreement was observed in the performance domain (.45). Based on the results of the pilot survey, several questions in the performance domain were rewritten to clarify the underlying concept of work performance.

For the purposes of the study, a personal communication device was defined as a wireless handheld device owned by an individual which can make and receive telephone calls or which provides a connection to the Internet via email, text messaging, videoconferencing, and social networking software. This definition includes cellular phones, mobile phones with app capabilities, and electronic tablet computers, but excludes desktop computers, pagers, or any company-provided device. Of the 825 respondents, 1.3 % (11/825) reported not owning a mobile phone or a comparable mobile communication device. For the purposes of reporting the results of the survey, the term mobile phone is therefore used as a general term to cover the multitude of mobile communication devices currently in use in hospitals.

In March 2014, a recruitment email containing a link to the previously validated 30-question survey was sent to the 10,978 members of the Academy of Medical-Surgical Nurses [8]. The Academy is a specialty nursing organization focused on medical-surgical nursing, with members based across the United States. Membership is open to anyone interested in medical-surgical nursing, including registered nurses, clinical nurse specialists, nurse practitioners, researchers, and administrators.

After excluding all respondents who did not meet the study criteria of having current full-time employment as a registered nurse in a hospital with an average of >5 hours a week of patient contact, the sample was divided into sub-groups for age and ethnicity. The subsets were examined to determine the representativeness of the sample relative to the US nursing workforce [7]. The probability that the proportions of various subgroups in the study sample were representative of the larger population of the US nursing workforce was calculated using a two-population z test to determine whether any weighting was necessary.

The results of the questions about self-reported and observed error rates were analyzed using a two-proportion z test (alpha=.05, two-tailed) to determine whether the differences between self-reported and observed outcomes were significant. All possible confounders which could potentially contribute to self-reported and observed performance errors were tested for significance, including age, race/ethnicity, length of time employed as a registered nurse, location of primary place of employment (inner city, rural, suburban, or urban), job title in primary nursing position, primary nursing role, number of hours per week spent caring for patients in an in-patient setting, US state of employment, type of primary place of employment (for profit, not for profit, or state/local government community hospital), and number of beds at primary place of employment.

Results

Overview

A total of 940 (8.56%, 940/10,978) respondents completed the Web-based questionnaire, and 825 (7.25%, 825/10,978) met the inclusion criteria for the study; having current full-time employment as a registered nurse in a hospital with an average of >5 hours a week of patient contact.

Demographics

Of the study sample, 48 (5.8%, 48/825) were male and 775 (93.9%, 775/825) were female. The age ranges were 20-30 years (9.3%, 77/825), 30-40 years (18.1%, 149/825), 40-50 years (23.9%, 197/825), 50-60 years (39.2%, 323/825), and >60 years (9.3%, 77/825). The results of the two-proportion *z* test indicated that sex and location of primary place of employment (urban/rural) were represented appropriately in the study sample. Respondents in the age groups <40 years old were underrepresented, while age groups >55 years old were overrepresented. Whites, American Indians/Alaskan natives, and Native Hawaiian or Pacific Islanders were underrepresented, while Hispanic and multiple/other ethnicities were overrepresented. Weighting the study sample data for age and race/ethnicity was determined to be undesirable because of the small sample sizes of several age groups and the inherent subjectivity of racial/ethnic groups.

Effects on Performance of Mobile Phone Use

Three survey questions assessed self-reported and witnessed performance decrements associated with the use of smartphones in the following areas (1) negative performance, (2) medical errors, and (3) missed clinical information. A significantly lower percentage of respondents self-reported mobile phone-related performance decrements (7.4%, 61/825) than reported witnessing mobile phone-related performance decrements in other nurses (70.9%, 584/825, $z=-26.6142$, $P<.001$, Fisher's exact test) (Table 1).

Significantly fewer respondents self-reported making medical errors (adverse effect of care, including a near-miss or a sentinel event) because of a mobile phone-related distraction (0.8%, 7/825) than reported witnessing such medical errors in other nurses (13.1%, 108/825, $z=-9.6798$, $P=.008$, Fisher's exact test) (Table 1). Likewise, significantly fewer respondents self-reported missing important clinical information because of mobile phone-related distractions (4%, 33/825) than reported witnessing other nurses missing important clinical information (29.9%, 246/825, $z=-13.9882$, $P<.001$, Fisher's exact test) (Table 1).

Table 1. Questionnaire responses (N=825).

Questions		Response, n (%)	
		Yes	No
Performance			
	Has mobile phone use ever negatively affected your performance as a nurse?	61 (7.4)	15 (1.8)
	Have you ever witnessed a nurse colleague's mobile phone use negatively affect his/her performance?	585 (70.9)	16 (1.9)
	z and Chi-square *P* values[a]	$z=-26.6142$, $P<.001$	
Medical error			
	Have you ever made a medical error (an adverse effect of care, including a near miss or a sentinel event) because you were distracted by your mobile phone?	7 (0.8)	34 (4.1)
	Have you ever witnessed a nurse make a medical error (defined as an adverse effect of care, including a near miss or sentinel event) because he/she was distracted by his/her mobile phone?	108 (13.1)	20 (2.4)
	z and Chi-square *P* values	$z=-9.6798$, $P=.008$	
Missed information			
	Do you think you have ever missed an important piece of clinical information because you were distracted by your mobile phone?	33 (4)	29 (3.5)
	Have you ever witnessed a nurse miss an important piece of clinical information because he/she was distracted by his/her mobile phone while working?	247 (29.9)	21 (2.5)
	z and Chi-square *P* values	$z=-13.9882$, $P<.001$	

[a]The *z* and Chi-square *P* values are for the difference between self-reported and observed errors in each case. *P* values were calculated using Fisher's exact test.

Serious Distraction

Of the 825 respondents, 351 believed that "smartphones can be a serious distraction during work"; always (13.0%, 107/825) or often (29.6%, 244/825). An additional 368 (44.6%, 368/825) believed that mobile phones were a serious distraction sometimes, and 82 thought that they were rarely (8.7%, 72/825) or never (1.2%, 10/825) a distraction. As well, a fraction, (2.9%, 24/825) did not answer the question.

Balance of Negative or Positive Effect on Patient Care

When asked, "On balance, do you think the use of personal smartphones by nurses while working has a more positive or negative effect on patient care?", the majority, 69.5% (573/825), said it has a more negative effect. Only 27.5% (227/825) said that the effect was more positive, and 25 respondents (3.0%, 25/825) did not answer the question.

Confounding Variables

Since all possible confounders were tested and none were significant, a multivariate analysis was not considered necessary. The proportion test was conducted under the assumption of normal approximation to the binomial distribution. The conditions for the normal approximation from the binomial distribution were satisfied. More advanced data analysis was not conducted because no confounding variables were significant.

Discussion

Principal Findings

An overwhelming majority of respondents answered that their work performance had never been negatively affected by their mobile phone use, with only 61 (7.4%, 61/825) reporting negative effects on performance, 7 (0.8%, 7/825) reporting that they had made a clinical error, and 33 (4%, 33/825) saying that they had missed an important piece of clinical information. However, far more said that they had witnessed a nurse colleague's work performance negatively affected due to the use of a mobile phone, with significant differences across all three categories. For example, 585 respondents (70.9%, 585/825) reported that they had seen a colleague's work performance negatively affected. More than 100 reported seeing someone else make a clinical error, and over one quarter claimed that they had seen a colleague miss a significant piece of clinical information.

Comparison With Prior Work

These results are consistent with findings from previous studies, including Katz-Sidlow et al's work on residents and attending physicians [4], and Smith et al's study of surgical technicians [5]. Both reports found significant discrepancies between self-reported and observed effects of mobile phone use on performance among the groups studied. There are several possible reasons for these discrepancies, including multiple observations of a smaller number of errors. However, this is unlikely in this study given the relatively small number of respondents coming from a wide geographical area, and the large differences between self-reported and observed outcomes. It is more likely that nurses may not be aware of their own mistakes while using their mobile phones or may not believe that they are distracted by them.

This conclusion is supported by work in other fields. For example, Lesch and Hancock [9] carried out a study to determine whether drivers were aware of their reduced driving ability while operating a mobile phone. They found that drivers were oblivious to their reduced driving ability caused by concurrent mobile phone use. Strayer et al [10] found that drivers described other drivers using their cell phones as driving poorly but reported that their own driving during mobile phone use remained normal, even when the results of driving performance tests showed otherwise. These results support our findings and suggest that there is a disconnect between self-reported and observed performance among respondents. Although respondents self-report low levels of performance decrements, the significantly higher level of reported witnessed performance decrements should be a cause for concern because it raises the possibility of substantial patient safety issues.

A recent study found that medical students believe that using mobile phones to support clinical work enables them to work more efficiently [11]. However, there were serious concerns raised in that study about the security of patient data transmitted via unsecured mobile phones, not least that current use may not be consistent with regulatory requirements. Similar to our results, this suggests that staff may be unable to accurately assess the risks of mobile phone use.

Limitations

This study has some limitations. First, the sample subgroups are not representative of the wider nursing population. Consideration was given to weighting the study sample data for race/ethnicity and age, but several points argued against it, such as the small sample sizes within several age groups and the inherent subjectivity of racial/ethnic groups. While the response rate was low relative to other online surveys, this may have been the result of the perceived sensitive nature of the subject, with respondents preferring not to admit that they had used their personal communication devices at work for non-work related activities. Holbrook et al [12] assessed whether lower response rates are associated with less unweighted demographic representativeness of a sample. By examining the results of 81 national surveys with response rates varying from 5%-54%, they found that surveys with much lower response rates were only minimally less accurate. As such, it was decided not to weight the current survey data, but to report the unweighted survey results with the recognition that the results, while valuable, may not be generalizable to the entire US registered nursing workforce.

Second, there may have been some selection bias. The usable response rate was quite low (7.25%, 60/825) as nurses may have been reluctant to participate because of the sensitive nature of the topic. This may have been a particular problem for any nurses who felt that their own performance had suffered as a result of personal mobile phone use and as such, the study overestimated the differences between the two groups. However, because these differences were quite large, this may not have affected the derived conclusions.

Finally, the results were based on respondents' recall and may not reflect current usage patterns. Real-time observations of mobile phone usage would provide a more accurate description of current usage patterns.

Conclusions

In response to concerns about mobile phone use, many hospitals are drawing up policies outlining the appropriate use of these devices at work. One approach is to allow workers to decide

how and when to use their devices at work. This presumes, however, that workers can accurately assess the risks associated with the use of their mobiles phones in a workplace environment and can appropriately modify their behavior. The results of this study indicate that registered nurses may express a disproportionately high confidence in their ability to manage the risk associated with the use of mobile phones at work relative to other registered nurses' performance, and may not be able to accurately assess when it is appropriate to use them and to modify their behavior accordingly.

Acknowledgments

We would like to thank the Academy of Medical-Surgical Nurses for their assistance in data collection. In addition, we would like to thank the Hanover Research Group for their help with the statistical analysis of the data in this paper. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' Contributions

Conceived and designed by DLM, SAL, and DL. DLM and SAL distributed the survey. Data analysis was performed by DLM, SAL, and DL. Analysis tools were contributed by DL. The manuscript was written by DLM, SAL, and DL.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Survey instrument used in the study.

[PDF File (Adobe PDF File), 226KB - humanfactors_v2i1e6_app1.pdf]

References

1. Ophir E, Nass C, Wagner AD. Cognitive control in media multitaskers. Proc Natl Acad Sci U S A 2009 Sep 15;106(37):15583-15587 [FREE Full text] [doi: 10.1073/pnas.0903620106] [Medline: 19706386]
2. Nasar JL, Troyer D. Pedestrian injuries due to mobile phone use in public places. Accid Anal Prev 2013 Aug;57:91-95. [doi: 10.1016/j.aap.2013.03.021] [Medline: 23644536]
3. Junco R, Cotten SR. No A 4 U: The relationship between multitasking and academic performance. Computers & Education 2012 Sep;59(2):505-514. [doi: 10.1016/j.compedu.2011.12.023]
4. Katz-Sidlow RJ, Ludwig A, Miller S, Sidlow R. Smartphone use during inpatient attending rounds: prevalence, patterns and potential for distraction. J Hosp Med 2012 Oct;7(8):595-599. [doi: 10.1002/jhm.1950] [Medline: 22744793]
5. Smith T, Darling E, Searles B. 2010 Survey on cell phone use while performing cardiopulmonary bypass. Perfusion 2011 Sep;26(5):375-380. [doi: 10.1177/0267659111409969] [Medline: 21593081]
6. Kiekkas P, Aretha D, Karga M, Karanikolas M. Self report may lead to underestimation of 'wrong dose' medication errors. Br J Clin Pharmacol 2009 Dec;68(6):963-964. [doi: 10.1111/j.1365-2125.2009.03530.x] [Medline: 20002092]
7. McBride DL, Levasseur SA, Li D. Development and validation of a Web-based survey on the use of personal communication devices by hospital registered nurses: pilot study. JMIR Res Protoc 2013;2(2):e50 [FREE Full text] [doi: 10.2196/resprot.2774] [Medline: 24280660]
8. Budden JS, Zhong EH, Moulton P, Cimiotti JP. Supplement: The National Council of State Boards of Nursing and the Forum of State Nursing Workforce Centers 2013 National Workforce Survey of Registered Nurses. J Nurs Regul 2013;4(2).
9. Lesch MF, Hancock PA. Driving performance during concurrent cell-phone use: are drivers aware of their performance decrements? Accid Anal Prev 2004 May;36(3):471-480. [doi: 10.1016/S0001-4575(03)00042-3] [Medline: 15003592]
10. Strayer DL, Watson JM, Drews FA. Cognitive distraction while multitasking in the automobile. Psychol Learn Motiv 2011;54:29-58. [doi: 10.1016/B978-0-12-385527-5.00002-4]
11. Tran K, Morra D, Lo V, Quan SD, Abrams H, Wu RC. Medical students and personal smartphones in the clinical environment: the impact on confidentiality of personal health information and professionalism. J Med Internet Res 2014;16(5):e132 [FREE Full text] [doi: 10.2196/jmir.3138] [Medline: 24855046]
12. Holbrook A, Krosnick J, Pfent A. The Causes and Consequences of Response Rates in Surveys by the News Media and Government Contractor Survey Research Firms. In: Advances in Telephone Survey Methodology (eds JM Lepkowski, C Tucker, JM Brick, ED d Leeuw, L Japec, PJ Lavrakas, MW Link and RL Sangster. Hoboken, NJ: John Wiley & Sons; 2007.

Edited by G Eysenbach; submitted 22.11.14; peer-reviewed by P Gill, K Tran; comments to author 16.12.14; revised version received 02.02.15; accepted 19.02.15; published 23.04.15.

Please cite as:
McBride D, LeVasseur SA, Li D
Nursing Performance and Mobile Phone Use: Are Nurses Aware of Their Performance Decrements?
JMIR Human Factors 2015;2(1):e6
URL: http://humanfactors.jmir.org/2015/1/e6/
doi:10.2196/humanfactors.4070
PMID:27026182

Original Paper

InformedTogether: Usability Evaluation of a Web-Based Decision Aid to Facilitate Shared Advance Care Planning for Severe Chronic Obstructive Pulmonary Disease

Lauren M Uhler[1], MPH; Rafael E Pérez Figueroa[2,3], MD, MPH; Mark Dickson[4], MA; Lauren McCullagh[1], MPH; Andre Kushniruk[5], PhD; Helen Monkman[5], MA; Holly O Witteman[6,7,8], PhD; Negin Hajizadeh[1], MD, MPH

[1]Department of Medicine, Hofstra North Shore LIJ School of Medicine, North Shore LIJ Health System, Manhasset, NY, United States

[2]Pediatrics and Population and Family Health, College of Physicians and Surgeons and Mailman School of Public Health, Columbia University Medical Center, New York, NY, United States

[3]International Family AIDS Global Health Program (IFAP), Columbia University Medical Center, New York, NY, United States

[4]Sitesteaders Development, Ypsilanti, MI, United States

[5]School of Health Information Science, University of Victoria, Victoria, BC, Canada

[6]Department of Family and Emergency Medicine, Laval University, Quebec City, QC, Canada

[7]Office of Education and Continuing Professional Development, Faculty of Medicine, Laval University, Quebec City, QC, Canada

[8]Research Centre of the CHU de Québec, Quebec City, QC, Canada

Corresponding Author:
Negin Hajizadeh, MD, MPH
Department of Medicine
Hofstra North Shore LIJ School of Medicine
North Shore LIJ Health System
300 Community Drive
Manhasset, NY,
United States
Phone: 1 1 516 465 5931
Fax: 1 1 516 465 5999
Email: nhajizadeh@nshs.edu

Abstract

Background: Advance care planning may help patients receive treatments that better align with their goals for care. We developed a Web-based decision aid called InformedTogether to facilitate shared advance care planning between chronic obstructive pulmonary disease (COPD) patients and their doctors.

Objective: Our objective was to assess the usability of the InformedTogether decision aid, including whether users could interact with the decision aid to engage in tasks required for shared decision making, whether users found the decision aid acceptable, and implications for redesign.

Methods: We conducted an observational study with 15 patients and 8 doctors at two ethnically and socioeconomically diverse outpatient clinics. Data included quantitative and qualitative observations of patients and doctors using the decision aid on tablet or laptop computers and data from semistructured interviews. Patients were shown the decision aid by a researcher acting as the doctor. Pulmonary doctors were observed using the decision aid independently and asked to think aloud (ie, verbalize their thoughts). A thematic analysis was implemented to explore key issues related to decision aid usability.

Results: Although patients and doctors found InformedTogether acceptable and would recommend that doctors use the decision aid with COPD patients, many patients had difficulty understanding the icon arrays that were used to communicate estimated prognoses and could not articulate the definitions of the two treatment choices—Full Code and Do Not Resuscitate (DNR). Minor usability problems regarding content, links, layout, and consistency were also identified and corresponding recommendations were outlined. In particular, participants suggested including more information about potential changes in quality of life resulting from the alternative advance directives. Some doctor participants thought the decision aid was too long and some thought it may cause nervousness among patients due to the topic area.

Conclusions: A decision aid for shared advance care planning for severe COPD was found acceptable to most COPD patients and their doctors. However, many patient participants did not demonstrate understanding of the treatment options or prognostic estimates. Many participants endorsed the use of the decision aid between doctors and their patients with COPD, although they desired more information about quality of life. The design must optimize comprehensibility, including revising the presentation of statistical information in the icon array, and feasibility of integration into clinical workflow, including shortening the decision aid.

(JMIR Human Factors 2015;2(1):e2) doi:10.2196/humanfactors.3842

KEYWORDS

usability testing; decision aid; shared decision making; COPD; advance care planning

Introduction

Chronic obstructive pulmonary disease (COPD) is a progressive disease affecting approximately 6.3% of adults (15 million) in the United States [1] and is the third leading cause of death in the United States [2]. As COPD advances, patients may experience COPD exacerbations—episodes in which their symptoms suddenly worsen, requiring hospitalization and a potential decision about whether to accept intubation. Patients in this situation who do not have advance directives often receive default invasive treatments, such as mechanical ventilation, which may not align with their goals of care [3]. Advance care planning (ACP) includes establishing advance directives and often involves discussions between the patient, family members, and outpatient clinicians [4]. Although most patients are open to discussing end-of-life issues, few have had such conversations with a doctor [5,6].

One strategy that has been advocated for improving patient-clinician communication is shared decision making [7]. Shared decision making is a process during which the clinician and patient work together to arrive at a decision that takes into consideration the patient's preferences. Decision aids are tools that can encourage informed, shared decision making by providing information to patients regarding their condition, available treatment options, and potential outcomes, and can also help them identify and communicate their preferences [8].

Decision aid experts have developed standards for the creation of high-quality decision aids [9], including following a systematic development process and performing iterative usability testing with patients and clinicians [10]. Usability testing is conducted with intended end users completing specific tasks using the decision aid prototype, while performance data are electronically captured and/or an observer records notes on what they do or say [11]. The purpose of usability testing is to identify specific problems that prevent users from reaching the goals of the decision aid—in this case, to be able to participate in shared decision making about advance directives on whether to receive invasive mechanical ventilation. Recommended solutions to usability problems are then incorporated into the decision aid during the iterative design process.

We have developed a Web-based decision aid called InformedTogether, which is designed to support shared advance care planning between severe COPD patients and their doctors. In this manuscript, we outline the results of usability testing of InformedTogether. Other decision aids about advance directives have been developed for COPD patients [12,13]. InformedTogether differs from these decision aids because it is intended to be used by the doctor and patient together during the clinic visit (ie, shared decision making) and then be made available to either party to access individually online. InformedTogether also includes personalized prognostic estimates using a published decision model based on the best available evidence of COPD outcomes [14,15]. Providing doctors with prognostic estimates may facilitate advance care planning [16] because uncertainty around a patient's illness trajectory has been identified as one reason doctors are reluctant to discuss end-of-life care planning [17,18]. The objective of this paper is to describe the usability of InformedTogether in terms of whether patients and doctors could use it to engage in tasks required for shared decision making about advance directives, whether they thought it was acceptable, and how the decision aid could be improved.

Methods

Development of the Decision Aid Prototype

The development and initial testing of InformedTogether was guided by the International Patient Decision Aid Standards (IPDAS) Collaboration criteria for quality decision aids [19]. The decision aid was designed to incorporate principles of shared decision making, including presenting patients with information about their treatment options and likely outcomes, presenting the risks and benefits of each option, and engaging the patient and physician in a conversation about the patient's preferences [7]. The decision presented is which advance directive to choose in the event of acute respiratory failure: (1) Full Code, which allows intubation for mechanical ventilation, or (2) Do Not Resuscitate (DNR), which does not allow invasive mechanical ventilation, but permits noninvasive ventilation with bilevel positive airway pressure (BiPAP) or continuous positive airway pressure (CPAP). InformedTogether was designed to be used on a Web-based platform, either on a tablet computer or on a desktop. Our previous research indicated that patients and doctors would be comfortable using the computer together during the clinic visit, and that patients would be open to their doctor using a decision aid [20].

Content of the Decision Aid Prototype

The decision aid allows clinicians to enter patient information including name, gender, and age. It then displays projected survival outcomes based on patient age and disease severity. The version used for this study calculated estimated outcomes

for a hypothetical patient aged 65 with severe COPD. Pages include a description of the goals of the decision aid, personalized survival estimates for Full Code versus DNR advance directives based on patients' age and severity of COPD, and suggested scripts for discussing the topics of prognosis and planning in case of a COPD exacerbation (see Multimedia Appendix 1 for decision aid screenshots).

Expert Consultation

Once the prototype was developed, we solicited feedback from experts not involved in the development and usability testing of the decision aid. We consulted experts in human factors engineering, health risk communication, and health care decision making to get feedback on interface design, and consulted palliative care experts on content and wording. We also consulted a patient advocate to provide feedback on content.

Usability Testing

Overview

Usability testing was conducted in two phases. The first phase focused on the icon array risk communication page, and the second phase tested interactions with the entire decision aid. An icon array—sometimes called a pictograph—is a graphical display of a number (usually 100 or 1000) of stick figures, circles, or other icons which represent individuals at risk of an event. The icons are shaded in one color to depict that they were affected by the event and unshaded to depict that they were not affected (see Figure 1). The icon array was created by a program developed at the Risk Science Center and Center for Bioethics and Social Sciences in Medicine, University of Michigan [21]. Usability testing focused on communication and understanding of treatment options, risks and benefits, and likely outcomes. Communication and understanding of patients' values and preferences, an important component of shared decision making, was not assessed as this feature was not included in the decision aid at the time of the studies. We also explored whether the decision aid would be feasible to implement in a real-world clinic setting by asking questions about acceptability.

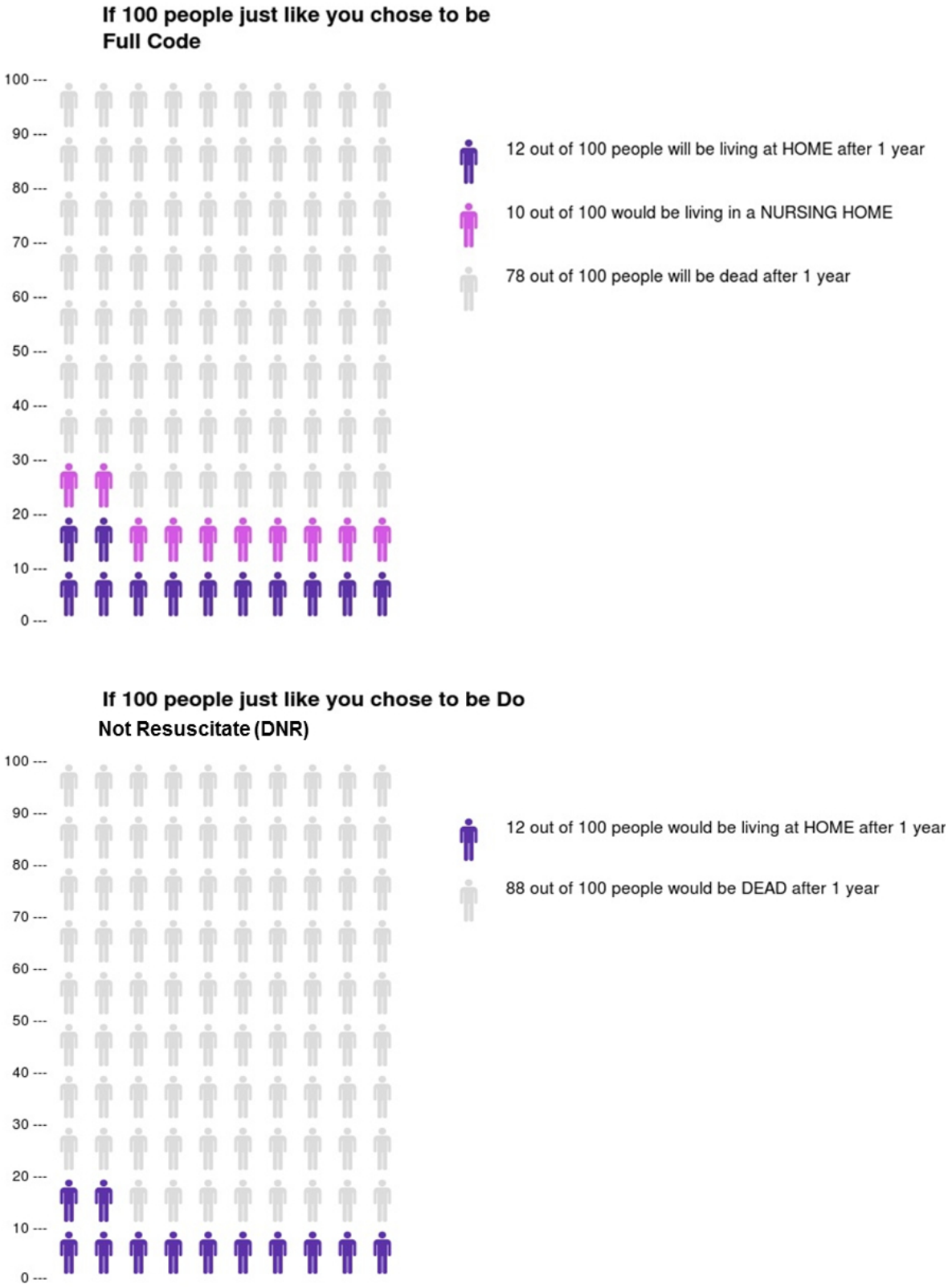
Phase 1: Patient Usability Testing of Icon Arrays in the Decision Aid

Phase 1 of usability testing was conducted over 1 week at the outpatient pulmonary clinic at Bellevue Hospital Center, a public hospital in New York City. The study protocol was approved by the NYU School of Medicine Institutional Review Board and by Bellevue Hospital Research Administration. Adult (18 years of age or older) English- or Spanish-speaking patients were approached in the waiting room and invited to participate. Interviews had two parts. The first part was designed to assess patients' and doctors' attitudes, knowledge, and preferences toward both shared decision making in general and shared end-of-life decision making. Those results are presented elsewhere [22]. The second part of the interview asked patient participants to view printed versions of the icon arrays (see Figure 1), and to explain what the pictures were showing and the meaning of Full Code and DNR in their own words. Participants were then asked about acceptability in terms of whether they would want their doctor to show them the icon arrays and whether they thought the icon array would help them make end-of-life plans or decisions. They were also asked about suggestions for improving the decision aid [23]. Interviews were audiotaped and transcribed.

Phase 2: Patient and Doctor Usability Testing of the Entire Decision Aid

Phase 2 of testing was conducted at a different center, with pulmonary rehabilitation patients and pulmonary doctors at Long Island Jewish Medical Center in New Hyde Park, NY. The protocol was approved by the NYU School of Medicine and the North Shore-Long Island Jewish Health System Institutional Review Boards. Adult (18 years of age or older) English-speaking patients with advanced-stage COPD who were receiving pulmonary rehabilitation were eligible for the study. All pulmonary doctors present during the day of the study were eligible. Each participant was shown the decision aid while being observed by a researcher. With patient participants, the researcher acted as the "doctor" and went through the decision aid on a tablet computer. Doctor participants used the decision aid on a laptop computer running Hypercam screen capture software (Hyperionics Technology LLC) and were asked to use the decision aid as if they were with a patient. Tasks analyzed included the following: (1) a click-through task, where users clicked through the decision aid, looking at the decision support materials while thinking aloud (ie, verbalizing their thoughts), (2) a graph interpretation task, where they were asked to respond to, and interpret, a graph, and (3) an icon array interpretation task, where they were asked to respond to, and interpret, an icon array. We then administered a brief, semistructured interview to all participants to assess their knowledge and understanding of the treatment choices presented in the decision aid, the acceptability of the decision aid with regard to the length, clarity, and amount of information, whether the participant would recommend use of the decision aid, and whether they had suggestions for improvements. Examples of questions to assess knowledge and understanding included "What did you think the overall message of the decision aid was?" and "In your own words, what is meant by 'Full Code'?" Questions to assess acceptability followed guidelines established by the Patient Decision Aids Research Group, [23] and included "How would you rate the amount of information in the decision aid?" and "Did the decision aid make you feel nervous or fearful?"

Figure 1. Icon arrays presented in phase 1 of usability testing. Likely outcomes 12 months after hospitalization for acute COPD exacerbation are shown for 100 hypothetical patients choosing either a Full Code or DNR advance directive.



Data Analysis Methods

Phase 1: Patient Usability Testing of Icon Arrays in the Decision Aid

We performed a thematic analysis of the transcribed audio recordings as recommended by Boyatzis [24]. The analytical process involved the following: (1) generating codes to be attached to similar quotes or topics across transcripts, (2) comparing and contrasting ideas related to the codes to create themes that fit the nature of the data, and (3) assessing the reliability of codes and themes. Data were analyzed keeping in mind key usability measures, such as understanding of the treatment choices (Full Code and DNR) and acceptability of the icon arrays. Qualitative data were coded and analyzed using NVivo qualitative data analysis software, version 10 (QSR International Pty Ltd).

Phase 2: Patient and Doctor Usability Testing of the Entire Decision Aid

For analysis of the click-through and think-aloud tasks, transcribed audio recordings of participant interactions with the decision aid were time-stamped (ie, video time codes were added into the file) and annotated for interesting user interactions and comments and to identify usability problems. The codes from the analysis were then summarized in terms of type and potential impact. For analysis of the interviews, closed-ended questions were summarized with descriptive statistics, and answers to open-ended questions were analyzed thematically as described above.

Results

Phase 1: Patient Usability Testing of Icon Arrays in the Decision Aid

Sample

Out of 52 eligible patients, 11 consented to participate in the study. The most common reason given for declining to participate was lack of time. Patient participants were mostly male (6/11, 55%), Hispanic or Black (10/11, 91%), and had a median age of 60 years, ranging from 23 to 73 (see Table 1). Most participants (7/11, 64%) had a high school education or less.

Understanding of Treatment Choices and Likely Outcomes

Participants understood that the chance of survival at 1 year was better with Full Code, and that survival was low regardless of advance directive. One participant said the following:

> *There's a better chance with the Full Code than the no resuscitation. The Full Code is with the tube right? It shows them that they could be longer, you know, you could help them out more with the Full Code. With the DNR you ain't got no chance. They ain't got no chance out of 100. (With Full Code) they got little chance out of 100.* [Male, 49, black/African American]

Some patients identified Full Code as *better* in terms of survival but had difficulty understanding the potential trade-off that more survivors of Full Code would be institutionalized within a nursing home. For instance, one participant started out saying "I would choose this one (Full Code) because then there would be a chance for longer life." After the interviewer pointed out that almost half of those who survive would be living in a nursing home, the participant said the following:

> *In that case, I wouldn't like that because I had to make a decision like that with my father and I decided to keep my father at my house and treat him at home. I would choose to live at home because I don't agree and have never agreed with being in a nursing home because the people who are in a nursing home die faster.* [Male, 60, Hispanic]

There was variable understanding of the meaning of the two advance directives. Although many participants responded in terms of whether the patient would get "the tube", many responded by describing survival outcomes for that advance directive. These answers did not describe what the advance directive meant in terms of treatments allowed. For example, one patient participant (male, 51, Hispanic) responded that DNR meant "...that they don't have life, that they die." Another participant (female, 35, Asian/Asian American) responded that Full Code meant "...you live longer...this is group of people, you're more statistically will prolong your life." Furthermore, several patient participants did not understand that choosing DNR could still mean patients could be treated with a breathing mask noninvasively.

Several participants misinterpreted the icon array in various ways. For instance, one participant thought that the numbers on the vertical axis represented age. Another considered *code* as referring to cardiac arrest and DNR to pertain only to the heart. When asked if patients who are DNR could get the breathing mask, the participant said the following:

> *They can get intubated as well because my concept of it is your heart has to stop or brain damage, things like that, but if your brain's dead, your brain's dead. If your heart stops that's when a resuscitation, but if you just coming in because you can't breathe I don't think that falls under the...As long as the heart is pumping, treatment can be given, but if that heart stops treatment cannot be given.* [Male, 34, black/African American]

Table 1. Participant characteristics.

Characteristics		Patients		Doctors
		Phase 1 (n=11)	Phase 2 (n=4)	Phase 2 (n=8)
Demographics				
Gender [a], n (%)				
	Female	5 (45)	1 (25)	N/A
	Male	6 (55)	3 (75)	N/A
Race/ethnicity, n (%)				
	Hispanic/Latino	7 (64)	0 (0)	0 (0)
	Black/African American	3 (27)	2 (50)	0 (0)
	White	0 (0)	2 (50)	3 (38)
	Asian/Asian American	1 (9)	0 (0)	4 (50)
Age in years, median (range)		60 (23-73)	72 (57-76)	33 (28-43)
Highest education level completed, n (%)				
	8^{th} grade or less	2 (18)	0 (0)	N/A
	9^{th} to 12^{th} grade	5 (45)	0 (0)	N/A
	Some college	2 (18)	0 (0)	N/A
	College degree	2 (18)	4 (100)	N/A
Years of training after residency, median (range)		N/A	N/A	2.5 (1-5)
Clinical characteristics				
Self-rated general health, n (%)				
	Excellent	1 (9)	0 (0)	N/A
	Very good	2 (18)	0 (0)	N/A
	Good	2 (18)	2 (50)	N/A
	Fair	4 (36)	1 (25)	N/A
	Poor	1 (9)	1 (25)	N/A

[a]Data on gender were not collected for doctors.

Acceptability

Overall, patient participants endorsed the use of the decision aid between patients and doctors. Participants articulated that death is “reality” and that patients need to know their options. Some patients said the decision aid should be used with family members present, and a few raised the concern that seeing the figures either caused them to be fearful or may cause fear in other patients. About one-third of patient participants said that the icon array would be helpful for patients in planning for end-of-life and about half said that they would like their doctors to show them a picture like the icon array (see Table 2).

Table 2. Results of patient interviews from phase 1 of usability testing.

Questions		Responses (n=11), n (%)
"Would these pictures be helpful for patients to see?", n (%)		
	Yes	4 (36)
	No	1 (9)
	Missing	6 (55)
"Would you like your doctor to show you something like this?", n (%)		
	Yes	5 (45)
	No	1 (9)
	Missing	5 (45)
"Do you think this would help you make plans or decisions about what you would want to happen at the end of your life?", n (%)		
	Yes	4 (36)
	No	1 (9)
	Missing	6 (55)

Suggestions for Improvements

When asked for suggestions for improving the decision aid, patient participants suggested increasing the size of the images and font to enable people to see them more clearly. Participants also wanted more information about outcomes such as the chance of successfully weaning from the ventilator. As one patient (male, 34, black/African American) stated, "It (the decision aid) should be something that forewarns you as far as what happens if you get intubated, as far as, like, there's a twenty percent chance it might never come out." A few participants who couldn't define DNR or Full Code requested the definitions to be included right next to the icon array images. For example, one participant stated the following:

> *I don't understand, like, what's really going on in the pictures and where the numbers come in, but yeah, I need some more information. This sheet, it's just saying Full Code and DNR, and also explain what Full Code means because...I think you know anyone that's not knowledgeable about these terminologies, it'd be good to break it down in simpler terms and explain exactly what these images are representing here.* [Male, 23, black/African American]

Phase 2: Patient and Doctor Usability Testing of the Entire Decision Aid

Sample

Four out of 5 (80%) patients and 8 out of 8 (100%) doctors that were approached consented to participate in the study. Table 1 provides demographic and clinical characteristics of the participants. Patient participants had a median age of 72 years (range 57 to 76), and rated their COPD as severe or very severe. All participants except for 1 (3/4, 75%) reported having an advance directive, but only 1 (1/4, 25%) participant reported having had an end-of-life discussion with their doctor. Doctor participants had a median age of 33 years (range 28 to 43), and all (8/8, 100%) reported having end-of-life discussions with their patients.

Understanding of Treatment Choices and Likely Outcomes

Patient Participants

Several patients responded to questions about what was meant by the terms Full Code and DNR by expressing their values and thoughts about that directive instead of by giving an objective definition or description of the term (see Table 3). For instance, one participant (male, 76, white) described DNR as the following: "If you have no chance of recovering, don't do all these things and you die soon anyway." Often, the patients did not express the level of understanding that was expected after using the decision aid. For example, one patient (male, 68, black/African American) defined Full Code as meaning "Let your wishes be known ahead of time." Change in knowledge could not be determined because there was no baseline assessment before viewing the decision aid.

When asked to interpret the survival curve and icon array, 2 out of 4 (50%) patients either interpreted some aspect of the icon arrays incorrectly or stated that they thought the icon arrays needed more clarity. Of note, 1 (1/2, 50%) of the patients (male, 57, black/African American), who later interpreted aspects of the icon array incorrectly, stated that he thought the decision aid was "...easy to see and understand," highlighting the necessity of using specific measures to assess understanding.

Table 3. Responses to questions assessing understanding of treatment options from phase 2 of usability testing.

Questions	Responses	Patient gender, age in years	Expresses understanding?, Yes/No
In your own words, what is meant by *Full Code* ?			
	When you are using oxygen or CPAP or the hose. Make the best of the situation and live.	Male, 57	No
	Code is when you're dying.	Male, 76	No
	Let your wishes be known ahead of time.	Male, 68	No
	CPR, meds, cardiac life support, intubate.	Female, 75	Yes
What is meant by *DNR* ?			
	If you have no chance of recovering, don't do all these things and you die soon anyway.	Male, 76	No
	Do not resuscitate.	Male, 68	No
	Do not resuscitate. Means don't do this, just do palliative care, leave me alone, make me comfortable even if it means drugs that may hasten my death.	Female, 75	Yes
	Gamble that you will pull through. Do not revive me. I'm giving up.	Male, 57	Yes

Doctor Participants

When asked to describe the survival curve and icon array as if they were with a patient, 3 doctors out of 8 (38%) stumbled over some aspect of describing the icon array, including one who described the shaded icons as representing people who did not survive the initial hospitalization. However, the icon array depicted outcomes in the 12 months after hospitalization. Another doctor described the icon arrays as showing 100 people with severe COPD who were "...divided into two groups," instead of as alternative treatment scenarios for the same people. Doctors' thoughts on the comprehensibility of the survival curve and icon array for their patients were mixed. Two doctors liked the survival curve better but thought the icon array may be more understandable to patients. Two doctors cautioned that the survival curve may not give patients all the information they need to make a decision, since it only provided information on survival, and not quality of life. One (1/8, 13%) doctor suggested that both figures be removed from the decision aid. One doctor stated that regardless of the graph, the physician should guide the patient in reviewing and interpreting it, saying, "A lot of people use numbers to guide their decision process. But (numbers) should be interpreted with caution. It's important to guide the patient in reviewing the graphs." Yet another doctor suggested including scripts to aid the doctor in describing the figures to patients.

Acceptability

Patient Participants

All patient participants (4/4, 100%) expressed interest in having their doctor use InformedTogether with them and stated that they would recommend their doctor use the decision aid with other COPD patients, suggesting demand for use of the decision aid. Patients recommended use of the decision aid in order to help provide information and options, and to achieve better decisions. For example, one patient (male, 76, white) said he would recommend that his doctor use the decision aid "...with all patients. Doctors should give patients the truth. The more information patients have, the better decision they can make." Most (3/4, 75%) patients reported not feeling nervous or fearful after using the decision aid. One patient (male, 57, black/African American) acknowledged that talking about death caused him "a little" nervousness or fearfulness, but still concluded that "...you've got to (have such discussions). The sooner the better. Don't beat around the bush."

Doctor Participants

All doctors (8/8, 100%) responded that they would recommend doctors use the decision aid with their patients. The most commonly cited reason was that it would help facilitate important discussions around end-of-life treatment options (4/8). One doctor (30, white) pointed out it would also facilitate earlier discussions, saying, "(We) need to have this conversation, but doctors are short on time. These conversations happen in the hospital, which is the wrong time." Out of 8 doctors, 3 (38%) mentioned concerns about time constraints during the clinic visit with regard to whether they would suggest use of the decision aid and its practicality for use within a regular clinic visit. For instance, one doctor (30, white) said the decision aid was "...probably too long for a regular clinic visit but the length is appropriate for this type of discussion." The time needed to use the decision aid ranged from approximately 15 to 20 minutes (10 to 15 minutes for the pages to be shared with the patient).

Suggestions for Improvement

Patient Participants

Patient participants expressed their desire for more information about treatment options such as lung transplant and BiPAP/CPAP, and for information about the quality-of-life implications of the different treatment options. For instance, patients wanted to know whether mechanical ventilation would be permanent and what would happen in the future with the choices presented.

Doctor Participants

Doctors also agreed with patient participants that the decision aid should provide more information about the implications of

the choices, including quality of life and functional status. Some doctors (3/8, 38%) mentioned the topic area as a reason for potential nervousness among their patients and several (4/8, 50%) said that the decision aid may make patients feel nervous or fearful, depending on the patient. One doctor (28, Asian/Asian American) said, “Some (patients) don’t like to discuss advance directives. It causes anxiety, but it depends on the patient.”

Analysis of the think-aloud and screen capture data collected yielded a number of other potential usability issues, which can be subsumed into content, consistency, layout, orientation, links, and feedback issues. For example, a layout problem experienced by several doctors was that they did not readily recognize that there was more information, such as the *next* and *back* buttons, below the visible screen and that they had to scroll to see all the information on the page. Usability issues and corresponding recommended changes to the decision aid are described in Table 4. Results are organized into short-term changes, which may be implemented easily without further research, and long-term changes, which require further research to understand how best to implement.

Table 4. Usability issues arising during phase 2 of usability testing and recommended decision aid changes.

Areas	Usability issues	Participants, age in years	Recommended changes
Short-term changes			
	Content		
	Lack of axes labels for figures.	Doctor, 43	Add axes to figure labels.
	Unclear which screens are meant to be shared with the patient.	Doctors, no age given, 28, 33	Add a note to doctors on the page before the start of screens meant to be shared with patients alerting them to share upcoming screens. Provide proper training/orientation prior to use.
	Pictures of patients depicting intubation do not show the tube clearly.	Patient, 75	Change the pictures used to show intubation.
	A page with definitions of advance directives was redundant.	Doctors, no age given, 31	Remove page with redundant definitions.
	Consistency		
	The figure for patients was described as "people out of one hundred", but the axis scale was written as proportions of 1.		Change axis to match wording.
	Back and *next* buttons are not located consistently but move depending on how much information is on the page.	Doctor, 33	Make location of buttons consistent across each page.
	Layout		
	Had to scroll to see all information—some users had difficultly realizing more information was below visible screen, such as the *back* and *next* buttons.	Doctors, 28, 33, 33, 33	Make graphic smaller to eliminate need for scrolling, reducing burden on working memory.
	Orientation		
	Some confusion, and took some time for doctors to get oriented to survival curve.	Doctors, no age given, 31	Add a suggested script for doctors to use when describing curve to patients. Provide proper training prior to use
	Unsure about what *Resources* link would lead to.	Doctor, 33	Make the name more descriptive.
	Links		
	Broken links on *Patient Resources* page led to webpage without information on advance directives.	Doctor, 33	Update links.
	Links should open in a new tab/window instead of replacing the decision aid in the window.	Doctor, 33	Reprogram so that links open in a new tab or window.
	Feedback/ links		
	Links on *Resources* page should change color once visited.	Doctor, 33	Change link color once it has been clicked on/visited.
	Feedback		
	Users unsure of how to finish using the decision aid and how to close it.	Doctor, 33	Change *I'm done* button to say *Exit*—put it in a more visible area.
	Unclear that last page is last page of decision aid.	Doctors, 33, 43	Add text box that appears after clicking to exit, making it clear the user has reached the end of the decision aid.
Long-term changes			
	Content		
	Icon arrays need more clarity.	Patient, 75	Scripts. Further usability testing and refinement.
	Lack of information regarding quality of life/functional status for patients in nursing homes, with intubation, and with BiPAP.	Doctors, 33, 30, 31, 43	Add more information about quality of life with various treatments and places of care.
	Probably too long for a regular clinic visit but the length is appropriate for this type of discussion.	Doctor, 30	Find potential areas to cut. Discussions with doctors about implementation.

Discussion

Principal Findings

Some patient and most doctor participants were able to use the decision aid to complete tasks required for shared decision making, although many patients had difficulty articulating the treatment options and understanding the icon arrays used to communicate estimated prognoses for each option. Many patient and doctor participants rated InformedTogether highly on measures of acceptability, including endorsing the use of the decision aid between doctors and their patients with severe COPD.

Usability testing provided insights into modifications that could improve usability of the decision aid, including minor issues related to content, layout, links, and feedback to the user. Several problems with comprehension were uncovered, especially with regard to understanding the icon array and the meanings of the two advance directives. Patient and doctor participants also suggested content to add, including quality-of-life implications of the advance directives. These content suggestions reflect important information needs to enable patients to engage in informed decision making using the decision aid.

Implications for Decision Aid Redesign

Next steps to continue development of InformedTogether will include adding information on quality of life and functional status when patients are intubated and when they are discharged to a nursing home. In addition, further testing of the revised icon array is needed to ensure the decision aid will be understandable to the majority of patients. Although participants in these phases of usability testing struggled to interpret the icon array correctly, other research has found that icon arrays are readily understandable to patients. In studies of alternative ways of presenting risk estimates in decision aids, icon arrays have been shown to be understandable to a majority of users and result in higher comprehension levels compared to other methods of displaying information [25-27]. Potential explanations for our results include poor labeling of the icon arrays, or differences in the demographic characteristics of the patient population. Another explanation could be low graphical literacy among our sample, which was not formally measured. Visual aids such as icon arrays have been shown to be especially helpful for communicating probabilities for people with low numeracy but with relatively high graphical literacy [28,29]. It may be that our population had low numeracy and low graphical literacy and, thus, icon arrays were not helpful in aiding comprehension of the statistical data. Other studies of icon arrays have found that up to 70% of individuals with low numeracy still answered risk understanding questions incorrectly, depending on their graphical literacy [29,30]. One study found that individuals with low graphical literacy had better comprehension when shown numbers instead of graphs [28]. Subsequent usability testing of InformedTogether will involve formal measures of participants' numeracy and graphical literacy [31]. We will also explore the possibility of having alternative presentations of risk data available, which can be tailored to participants' numerical or graphical literacy, or can be chosen by patient users themselves. IPDAS standards recommend allowing the patient to choose how they view probabilities—either in numbers, words, or figures [19]. To optimize patient understanding and usefulness of the decision aid to patients with low numeracy and graphical literacy, it may be important to train health care providers on how to best communicate the risk information in the decision aid to ensure that it is understood by patients with different literacy levels [32]. Once revisions have been made to the icon array content based on the results of these phases of testing, we will again evaluate understanding and compare comprehension of the icon array with that of alternative forms of risk data presentation.

We will also need to improve the communication of information about the different advance directives and the questions used to test comprehension. Several patients offered their opinion or feelings about each directive rather than giving a definition in terms of allowable treatments. Less equivocal results on patients' understanding of advance directives may have been achieved by using closed-ended questions instead of, or in addition to, open-ended questions. For example, participants could be asked "Does a patient choosing a Full Code directive wish to allow intubation?" Future testing will include closed-ended questions as well as a baseline assessment of knowledge.

Revisions must also focus on allowing the decision aid to fit within the workflow in terms of time constraints. In addition, the decision aid must also support doctors' attempts to establish rapport with their patients. In future testing, we will specifically elicit users' attitudes about the effect of the decision aid on doctor-patient rapport. Finally, one component of shared decision making that had not yet been developed in this version of InformedTogether was patient preference elicitation and the communication and comprehension of these preferences. We recognize that including methods for values clarification and communication is an IPDAS criterion and an essential component of patient decision aids [19]. However, we intended to first test whether patients understood the information in this prototype before adding values clarification. Future versions will incorporate this central feature of shared decision making and test this feature for usability. Other relevant IPDAS criteria that were not met due to the early stage of the prototype include providing structured guidance for deliberation and discussion of the decision with others, complete information on the evidence used in the decision aid, including references and the quality of scientific evidence, and information about the developers and their conflicts of interest.

Results in the Context of Other Studies of COPD Decision Aids

Other investigators have tested the communication of statistical information to COPD patients in usability testing of decision aids related to mechanical ventilation and other aspects of COPD treatment. In the evaluation of their decision aid for COPD, Wilson et al reported that prior to the evaluation study, their decision aid had undergone several revisions, although they did not describe iterative usability testing in detail [33]. Their audio booklet decision aid for mechanical ventilation was tested by severe COPD patients, about half of whom had a high school education or higher. Similar to the present findings, the majority of users (88%) reported that the decision aid was "not at all"

difficult to understand, however, in analysis of open-ended questions to ascertain understanding, almost one-quarter of participants showed an inadequate comprehension of the options presented in the decision aid [33]. These results highlight the usefulness of a mixed-methods approach for a deeper understanding of results. In our studies, only 3 patients out of 15 (20%) reported that the decision aid made him or her feel nervous or fearful, but in Wilson et al's study, almost half the participants (15/33, 45%) found the decision aid to be at least a little upsetting. These results underline that it is important for doctors to assess their patients' readiness for using the decision aid in order to address any nervousness or discomfort.

For their computer-based decision aid regarding inhaled steroid therapy for COPD, Akl et al described usability testing consisting of interviews with 7 COPD patients followed by pilot-testing with 8 COPD patients [34]. Similar to the findings from our studies, the main modifications based on usability testing included changing the presentation of statistical information and the amount of information. Results from their pilot test showed improvement in knowledge from baseline, but users did not rate the clarity of statistical information highly [34]. The best ways of presenting statistical information to patients, including risk information, is an ongoing challenge that deserves further study.

Limitations and Strengths

As described above, one limitation of our study was the way comprehension of treatment options was measured. Inclusion of closed-ended knowledge questions, as well as a baseline measure of knowledge, would have helped clarify what effect, if any, the decision aid had on understanding of treatment options. A further potential limitation was that doctor participants may not have been given adequate time to orient themselves to the decision aid before testing. Although they were provided with 5 minutes to look through the pages before recording began, the decision aid is meant to be used by doctors who have had formal training with the tool and the results of our testing may, therefore, not fully reflect the intended use. For example, doctors' errors in interpretation of the icon arrays may have been avoided if they had been given the orientation and training in use of the decision aid that is planned for when the tool is disseminated.

We did not include testing of the decision aid during a clinic consultation for these early rounds of usability testing. This paper reports initial usability testing of a decision aid prototype which is being developed further before conducting feasibility testing in the clinical setting. Feasibility testing is planned to measure whether the decision aid promotes shared decision making and informs patients, and will measure outcomes such as demand and implementation, as well as patients' preparation for decision making, motivation to make advance care plans, confidence in decisions, and patient-doctor communication. Feasibility testing will also measure knowledge, comprehension, and acceptability. It is unknown whether testing within the clinic would show similar levels of comprehension or acceptability, however, our study was designed to enable critical revisions to the prototype before testing in the clinical setting.

Patient participants in phase 2 of usability testing were well educated (ie, all had college degrees) and their responses may not represent those of patients with broad educational backgrounds. However, patient participants in phase 1 had a larger range of education attainment, with most having completed 12th grade or less. Another limitation related to our sample was the small sample size, however, experts recommend 6 to 12 participants to detect the majority of usability problems [35]. While small sample sizes are adequate to uncover most usability problems, the small sample size may have affected the results regarding acceptability. For example, patients who agreed to participate in the study may have been more open to discussing end-of-life issues, and thus more likely to recommend use of a decision aid about end-of-life decisions. In phase 1, most eligible patients declined to participate. Although we did not systematically gather data about decliners, the reason most often given for declining was lack of time. We recruited patients from the waiting room of a clinic where it is common for patients to experience wait times of 2 hours or more, and asked them to stay for approximately 30 minutes after their appointment to participate. Those who participated may have been less likely to have commitments, such as full-time jobs or child or eldercare responsibilities, that would have prevented them from being able to stay longer at the clinic. The younger average age of the sample of doctor participants may have influenced the results if, for example, younger doctors tend to be more comfortable using computers or less worried about the impact of a computer-based tool on the patient-doctor relationship. We did not, however, measure doctors' comfort with, or preferences for, using computer-based tools with their patients. Subsequent usability testing, which is ongoing, involves doctors with varying levels of seniority. We also did not test the effectiveness of the decision aid in terms of constructs related to the decision-making process and decision quality, such as preparation for decision making [36].

Our study has several strengths, including the formal evaluation of screen capture recordings to identify usability problems. In addition, our sample included intended end users—doctors who treat COPD patients and COPD patients from two different clinic locations. Furthermore, we included patients from racially and socioeconomically diverse backgrounds. Our results provide directions for further refinement and development of the decision aid to ensure that it is usable and useful to both patients and doctors.

Acknowledgments

The authors would like to thank the US Agency for Healthcare Research and Quality for providing a grant (K12 HS019473-01) to the NYU School of Medicine, which partially supported this work. We also thank the National Institute on Drug Abuse (NIDA) for funding Dr Pérez Figueroa (R25 DA026401), a fellow of the Interdisciplinary Research Training Institute (IRTI) on Hispanic

Drug Abuse. We also thank the Fonds de Recherche du Québec - Santé for funding Dr Witteman with a Research Scholar Junior 1 Career Development Award.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Screenshots of the InformedTogether decision aid.

[PDF File (Adobe PDF File), 1MB - humanfactors_v2i1e2_app1.pdf]

References

1. Centers for Disease Control and Prevention (CDC). Chronic obstructive pulmonary disease among adults--United States, 2011. MMWR Morb Mortal Wkly Rep 2012 Nov 23;61(46):938-943 [FREE Full text] [Medline: 23169314]
2. National Heart, Lung, and Blood Institute. What is COPD? URL: http://www.nhlbi.nih.gov/health/health-topics/topics/copd/ [accessed 2015-02-10] [WebCite Cache ID 6WF20FElJ]
3. Lynn J, Teno JM, Phillips RS, Wu AW, Desbiens N, Harrold J, et al. Perceptions by family members of the dying experience of older and seriously ill patients. SUPPORT Investigators. Study to understand prognoses and preferences for outcomes and risks of treatments. Ann Intern Med 1997 Jan 15;126(2):97-106. [Medline: 9005760]
4. Mori M, Ellison D, Ashikaga T, McVeigh U, Ramsay A, Ades S. In-advance end-of-life discussions and the quality of inpatient end-of-life care: a pilot study in bereaved primary caregivers of advanced cancer patients. Support Care Cancer 2013 Feb;21(2):629-636. [doi: 10.1007/s00520-012-1581-x] [Medline: 22933133]
5. Heyland DK, Tranmer J, O'Callaghan CJ, Gafni A. The seriously ill hospitalized patient: preferred role in end-of-life decision making? J Crit Care 2003 Mar;18(1):3-10. [doi: 10.1053/jcrc.2003.YJCRC2] [Medline: 12640606]
6. Knauft E, Nielsen EL, Engelberg RA, Patrick DL, Curtis JR. Barriers and facilitators to end-of-life care communication for patients with COPD. Chest 2005 Jun;127(6):2188-2196. [doi: 10.1378/chest.127.6.2188] [Medline: 15947336]
7. Elwyn G, Frosch D, Thomson R, Joseph-Williams N, Lloyd A, Kinnersley P, et al. Shared decision making: a model for clinical practice. J Gen Intern Med 2012 Oct;27(10):1361-1367 [FREE Full text] [doi: 10.1007/s11606-012-2077-6] [Medline: 22618581]
8. Barry MJ, Edgman-Levitan S. Shared decision making--pinnacle of patient-centered care. N Engl J Med 2012 Mar 1;366(9):780-781. [doi: 10.1056/NEJMp1109283] [Medline: 22375967]
9. Elwyn G, O'Connor A, Stacey D, Volk R, Edwards A, Coulter A, International Patient Decision Aids Standards (IPDAS) Collaboration. Developing a quality criteria framework for patient decision aids: online international Delphi consensus process. BMJ 2006 Aug 26;333(7565):417 [FREE Full text] [doi: 10.1136/bmj.38926.629329.AE] [Medline: 16908462]
10. Coulter A, Stilwell D, Kryworuchko J, Mullen PD, Ng CJ, van der Weijden T. A systematic development process for patient decision aids. BMC Med Inform Decis Mak 2013;13 Suppl 2:S2. [doi: 10.1186/1472-6947-13-S2-S2] [Medline: 24625093]
11. Usability. Washington, DC: US Department of Health & Human Services Usability testing URL: http://www.usability.gov/how-to-and-tools/methods/usability-testing.html [accessed 2014-09-02] [WebCite Cache ID 6SIC3X0bN]
12. Dales RE, O'Connor A, Hebert P, Sullivan K, McKim D, Llewellyn-Thomas H. Intubation and mechanical ventilation for COPD: development of an instrument to elicit patient preferences. Chest 1999 Sep;116(3):792-800 [FREE Full text] [Medline: 10492288]
13. Hebert P, O'Connor A, Aaron S, Wilson K, Dales R, Fiset V, et al. Making Choices: The Use of Intubation and Mechanical Ventilation for Severe Chronic Obstructive Pulmonary Disease. 2008. URL: http://decisionaid.ohri.ca/docs/das/COPD.pdf [accessed 2015-02-10] [WebCite Cache ID 6WF2GH3Uq]
14. Hajizadeh N, Crothers K, Braithwaite RS. Informing shared decisions about advance directives for patients with severe chronic obstructive pulmonary disease: a modeling approach. Value Health 2012;15(2):357-366. [doi: 10.1016/j.jval.2011.10.015] [Medline: 22433768]
15. Hajizadeh N, Crothers K, Braithwaite RS. Using modeling to inform patient-centered care choices at the end of life. J Comp Eff Res 2013 Sep;2(5):497-508 [FREE Full text] [doi: 10.2217/cer.13.53] [Medline: 24236746]
16. Murray SA, Kendall M, Boyd K, Sheikh A. Illness trajectories and palliative care. BMJ 2005 Apr 30;330(7498):1007-1011 [FREE Full text] [doi: 10.1136/bmj.330.7498.1007] [Medline: 15860828]
17. Crawford A. Respiratory practitioners' experience of end-of-life discussions in COPD. Br J Nurs 2010;19(18):1164-1169. [doi: 10.12968/bjon.2010.19.18.79049] [Medline: 20948471]
18. Murray SA, Boyd K, Sheikh A. Palliative care in chronic illness. BMJ 2005 Mar 19;330(7492):611-612 [FREE Full text] [doi: 10.1136/bmj.330.7492.611] [Medline: 15774965]
19. International Patient Decision Aid Standards (IPDAS) Collaboration. IPDAS 2005: Criteria for Judging the Quality of Patient Decision Aids.: IPDAS URL: http://ipdas.ohri.ca/IPDAS_checklist.pdf [accessed 2015-02-10] [WebCite Cache ID 6WF2NJ25l]

20. Hajizadeh N, Perez Figueroa RE, Uhler LM, Chiou E, Perchonok JE, Montague E. Identifying design considerations for a shared decision aid for use at the point of outpatient clinical care: An ethnographic study at an inner city clinic. J Particip Med 2013 Mar 6;5 [FREE Full text] [Medline: 24748995]
21. Icon Array. Ann Arbor, MI: Risk Science Center and Center for Bioethics and Social Sciences in Medicine, University of Michigan URL: http://www.iconarray.com/ [accessed 2015-02-05] [WebCite Cache ID 6W7TDqT5r]
22. Hajizadeh N, Uhler LM, Pérez Figueroa RE. Understanding patients' and doctors' attitudes about shared decision making for advance care planning. Health Expect 2014 Oct 22. [doi: 10.1111/hex.12285] [Medline: 25336141]
23. O'Connor AM, Cranney A. User Manual-Acceptability. Ottawa, ON: Ottawa Hospital Research Institute; 2002. URL: https://decisionaid.ohri.ca/docs/develop/User_Manuals/UM_Acceptability.pdf [accessed 2015-02-10] [WebCite Cache ID 6WF3bGmxm]
24. Boyatzis RE. Transforming Qualitative Information: Thematic Analysis and Code Development. Thousand Oaks, CA: SAGE Publications, Inc; 1998.
25. Zikmund-Fisher BJ, Ubel PA, Smith DM, Derry HA, McClure JB, Stark A, et al. Communicating side effect risks in a tamoxifen prophylaxis decision aid: the debiasing influence of pictographs. Patient Educ Couns 2008 Nov;73(2):209-214 [FREE Full text] [doi: 10.1016/j.pec.2008.05.010] [Medline: 18602242]
26. Price M, Cameron R, Butow P. Communicating risk information: the influence of graphical display format on quantitative information perception-Accuracy, comprehension and preferences. Patient Educ Couns 2007 Dec;69(1-3):121-128. [doi: 10.1016/j.pec.2007.08.006] [Medline: 17905553]
27. Fagerlin A, Wang C, Ubel PA. Reducing the influence of anecdotal reasoning on people's health care decisions: is a picture worth a thousand statistics? Med Decis Making 2005;25(4):398-405. [doi: 10.1177/0272989X05278931] [Medline: 16061891]
28. Gaissmaier W, Wegwarth O, Skopec D, Müller AS, Broschinski S, Politi MC. Numbers can be worth a thousand pictures: individual differences in understanding graphical and numerical representations of health-related information. Health Psychol 2012 May;31(3):286-296. [doi: 10.1037/a0024850] [Medline: 21842998]
29. Garcia-Retamero R, Galesic M. Who profits from visual aids: overcoming challenges in people's understanding of risks [corrected]. Soc Sci Med 2010 Apr;70(7):1019-1025. [doi: 10.1016/j.socscimed.2009.11.031] [Medline: 20116159]
30. Garcia-Retamero R, Galesic M. Communicating treatment risk reduction to people with low numeracy skills: a cross-cultural comparison. Am J Public Health 2009 Dec;99(12):2196-2202. [doi: 10.2105/AJPH.2009.160234] [Medline: 19833983]
31. Galesic M, Garcia-Retamero R. Graph literacy: a cross-cultural comparison. Med Decis Making 2011 Jun;31(3):444-457. [doi: 10.1177/0272989X10373805] [Medline: 20671213]
32. Trevena LJ, Zikmund-Fisher BJ, Edwards A, Gaissmaier W, Galesic M, Han PK, et al. Presenting quantitative information about decision outcomes: a risk communication primer for patient decision aid developers. BMC Med Inform Decis Mak 2013;13 Suppl 2:S7. [doi: 10.1186/1472-6947-13-S2-S7] [Medline: 24625237]
33. Wilson KG, Aaron SD, Vandemheen KL, Hébert PC, McKim DA, Fiset V, et al. Evaluation of a decision aid for making choices about intubation and mechanical ventilation in chronic obstructive pulmonary disease. Patient Educ Couns 2005 Apr;57(1):88-95. [doi: 10.1016/j.pec.2004.04.004] [Medline: 15797156]
34. Akl EA, Grant BJ, Guyatt GH, Montori VM, Schünemann HJ. A decision aid for COPD patients considering inhaled steroid therapy: development and before and after pilot testing. BMC Med Inform Decis Mak 2007;7:12 [FREE Full text] [doi: 10.1186/1472-6947-7-12] [Medline: 17504536]
35. Dumas JS, Redish JC. A Practical Guide to Usability Testing. Revised edition. Portland, OR: Intellect Books; 1999.
36. Sepucha KR, Borkhoff CM, Lally J, Levin CA, Matlock DD, Ng CJ, et al. Establishing the effectiveness of patient decision aids: key constructs and measurement instruments. BMC Med Inform Decis Mak 2013;13 Suppl 2:S12. [doi: 10.1186/1472-6947-13-S2-S12] [Medline: 24625035]

Abbreviations

ACP: advance care planning
BiPAP: bilevel positive airway pressure
COPD: chronic obstructive pulmonary disease
CPAP: continuous positive airway pressure
DNR: Do Not Resuscitate
IPDAS: International Patient Decision Aid Standards
IRTI: Interdisciplinary Research Training Institute
NIDA: National Institute on Drug Abuse

Edited by G Eysenbach; submitted 04.09.14; peer-reviewed by L Kriston, J Dirmaier; comments to author 15.12.14; revised version received 21.01.15; accepted 21.01.15; published 25.02.15.

Please cite as:
Uhler LM, Pérez Figueroa RE, Dickson M, McCullagh L, Kushniruk A, Monkman H, Witteman HO, Hajizadeh N
InformedTogether: Usability Evaluation of a Web-Based Decision Aid to Facilitate Shared Advance Care Planning for Severe Chronic Obstructive Pulmonary Disease
JMIR Human Factors 2015;2(1):e2
URL: http://humanfactors.jmir.org/2015/1/e2/
doi:10.2196/humanfactors.3842
PMID:27025896

Original Paper

Knowledge and Utilization of Computers Among Health Professionals in a Developing Country: A Cross-Sectional Study

Kalid Alwan[1*], MPH; Tadesse Awoke[2*], MSc; Binyam Tilahun[3*], MSc, MPH

[1]Harar Health Science College, Harar, Ethiopia

[2]Department of Epidemiology and Biostatistics, University of Gondar, Gondar, Ethiopia

[3]Institute of Medical Informatics, University of Münster, Münster, Germany

[*]all authors contributed equally

Corresponding Author:
Binyam Tilahun, MSc, MPH
Institute of Medical Informatics
University of Münster
Albert-Schweitzer-Campus 1, Gebäude A11
MÃ¼nster,
Germany
Phone: 49 251 83 58368
Fax: 49 251 83 52259
Email: binyam.tilahun@uni-muenster.de

Abstract

Background: Incorporation of information communication technology in health care has gained wide acceptance in the last two decades. Developing countries are also incorporating information communication technology into the health system including the implementation of electronic medical records in major hospitals and the use of mobile health in rural community-based health interventions. However, the literature on the level of knowledge and utilization of information communication technology by health professionals in those settings is scarce for proper implementation planning.

Objective: The objective of this study is to assess knowledge, computer utilization, and associated factors among health professionals in hospitals and health institutions in Ethiopia.

Methods: A quantitative cross-sectional study was conducted on 554 health professionals working in 7 hospitals, 19 primary health centers, and 10 private clinics in the Harari region of Ethiopia. Data were collected using a semi-structured, self-administered, and pre-tested questionnaire. Descriptive and logistic regression techniques using SPSS version 16.0 (IBM Corporation) were applied to determine the level of knowledge and identify determinants of utilization of information communication technology.

Results: Out of 554 participants, 482 (87.0%) of them responded to the questionnaire. Among them, 90 (18.7%) demonstrated good knowledge of computers while 142 (29.5%) demonstrated good utilization habits. Health professionals who work in the primary health centers were found to have lower knowledge (3.4%) and utilization (18.4%). Age (adjusted odds ratio [AOR]=3.06, 95% CI 0.57-5.37), field of study (AOR=3.08, 95% CI 1.65-5.73), level of education (AOR=2.78, 95% CI 1.43-5.40), and previous computer training participation (AOR=3.65, 95% CI 1.62-8.21) were found to be significantly associated with computer utilization habits of health professionals.

Conclusions: Computer knowledge and utilization habits of health professionals, especially those who work in primary health centers, were found to be low. Providing trainings and continuous follow-up are necessary measures to increase the likelihood of the success of implemented eHealth systems in those settings.

(JMIR Human Factors 2015;2(1):e4) doi:10.2196/humanfactors.4184

KEYWORDS

computer literacy; health professionals; eHealth success; Ethiopia

Introduction

Background

The use of information communication technology in health care is not merely about technology but a means to solve the critical data management and clinical communication challenges in health care organizations, especially in developing countries [1]. Given the high burden of disease and the low number of skilled personnel, eHealth is believed to improve health care by strengthening the health system, supporting delivery of care, and improving communication among different health care organizations and professionals [2,3]. Incorporation of information communication technology in developing countries has gained wide acceptance in the last several decades with different success stories in different sectors, especially in the business sector [4]. However, when compared with other sectors, only a limited application of information technology advancements is seen in health care organizations [5,6].

Recently there has been an increase in the implementation of eHealth applications in developing countries that includes telehealth, mobile health applications, electronic medical records, and health information management systems [7]. However, most implementations remain in the pilot phase because of different technical and personnel issues [8]. Most evaluations and case studies from previous implementations in those settings report that infrastructural challenges and the existing skill levels of health professionals are the most common obstacles to the success of implemented eHealth systems [8,9,10]. However, the literature on the level of knowledge and utilization of health professionals and their current exposure in information communication technology use is scarce.

For Ethiopia, with its population of approximately 80 million people, poor health system, and severe shortage of health professionals, incorporation of eHealth to the different sectors of the system is regarded as the only way to achieve the country's goal of universal health coverage by 2020. For that, the government is currently implementing different eHealth initiatives, and the Health Sector Development Plan IV [11] strategy is in progress to transform health services into a cost effective and efficient system. The Ministry of Health of Ethiopia is also drafting a new national eHealth strategy [12] following the recently published World Health Organization guideline [13] on eHealth strategy development. To ensure sustainability, the country is also teaching health informatics professionals [14] who support different eHealth implementation initiatives in the country.

Statement of the Problem

With the new initiatives in Ethiopia, expanded implementation of eHealth is expected in the coming years, but these systems must be used effectively to meet objectives; this is entirely dependent on health professionals' use of eHealth in their daily tasks. Studies in similar settings show that that lack of basic knowledge of computers and software on the part of health professionals is a main factor in failure of eHealth systems [5,15-18]. Therefore, before costly implementation, it is necessary to know the current knowledge and utilization habits of health professionals so that effective prior planning can take place. To our knowledge, there is little existing evidence in primary care and hospital contexts in developing countries. This study aims to fill this gap.

Objectives

The goal of this study is to assess the current levels of knowledge and utilization of computers among health professionals and identify factors affecting utilization. The outcome of this research will help evidence-based planning and implementation of eHealth in Ethiopia and generate additional insight on the topic for further development of health systems in other developing countries.

Methods

Overview

Institution-based quantitative cross-sectional research was conducted in 7 hospitals, 19 primary health centers, and 10 private clinics which are on the frontline to implement different eHealth applications in the coming year. All health professionals working at these health institutions were included in the study. There were 621 health professionals working at those institutions; all except those on annual and sick leave were included in the study.

A pre-tested, self-administered questionnaire, adapted from a previous study [19](Multimedia Appendix 1), was used to collect data on sociodemographic characteristics and computer knowledge and utilization by health professionals. The questionnaire was prepared in English. The data collection was facilitated by eight information technology diploma holders, and supervision was done by the principal investigator.

In this study, health professionals were defined as those employees with at least a diploma certificate in the health professions who are practicing clinical service in the study settings. Computer knowledge was defined as a basic understanding about computers and how to use them. It involves knowing hardware and software, what a computer virus is, and how to manage files and use basic computer applications like a computer network and the Internet. Twenty questions were used to assess computer knowledge. Utilization of computers is a basic skill and involves use of the computer and Internet; managing and storing files; and retrieving, analyzing, and presenting the data on hand. Fifteen questions were used to assess health worker computer utilization habits.

Both knowledge and utilization of computers among health professionals were classified after adopting a cut of value from the Nigerian study in 2004 on the same topic [20]. Those scoring 80% or above on the knowledge test were rated as having good computer knowledge; those scoring below 80% were rated as having poor computer knowledge**.** Those study subjects who scored 60% or above on the utilization test were rated as having good computer utilization, while those scoring below 60% were rated as having poor computer utilization.

Data were entered using Epi Info then exported to SPSS package version 16 (IBM Corporation) for analysis. Frequencies and cross tabulations were used for the descriptive analysis of the data. Associations between participant's characteristics and

knowledge and utilization of computer were analyzed using binary logistic regression.

Ethics Statement

The ethical clearance committee of the University of Gondar College of Medicine and Health Sciences through the School of Public Health approved this study. Data were collected after getting permission and clearance from the ethical clearance committee of the Harari Regional Health Bureau. Written consent was obtained from each respondent on a form attached to the questionnaire.

Results

Sociodemographic Characteristics

There were 554 public health professionals who participated in this study. Among them, 482 (87.0%) correctly filled out and returned the questionnaire. The median age of respondents was 25 years; 52.0% (251/482) were male. The majority of participants (311/482, 65.0%) were nurses while 20.7% (100/482) were pharmacists and laboratory technicians. Most respondents (364/482, 75.5%) had received at least some kind of basic computer training in the past.

Table 1. Sociodemographic characteristics of respondents.

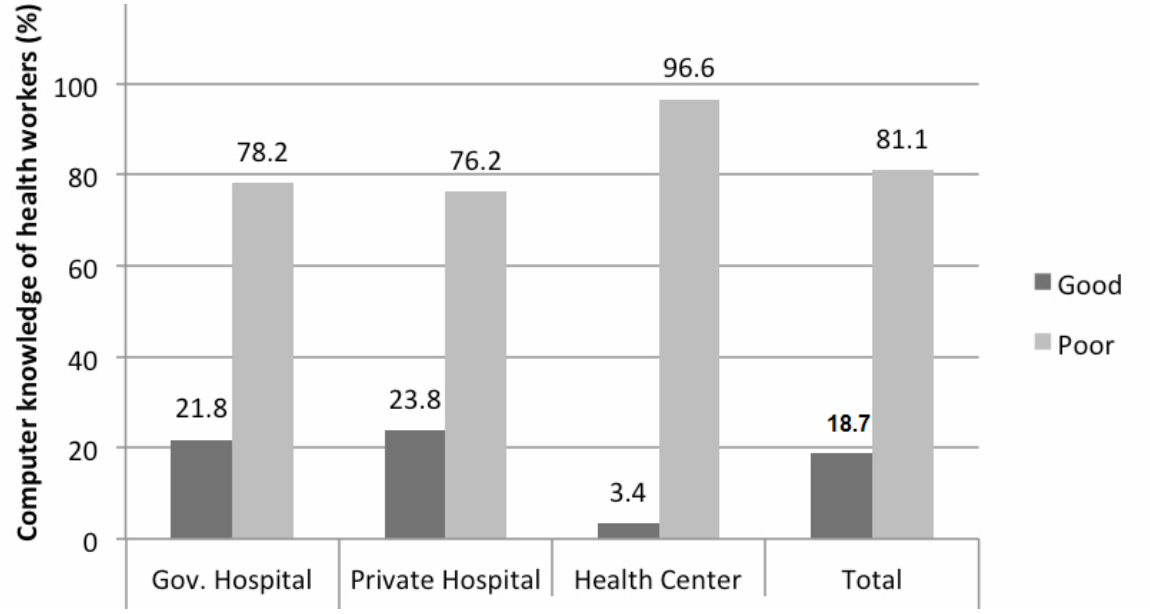
Predictor variables of respondents		n (%)
Age, years		
	≤25	174 (36.1)
	26-30	118 (24.5)
	31-35	71 (14.7)
	≥36	119 (24.7)
Sex		
	Male	251 (52.1)
	Female	231 (47.1)
Profession		
	Medical doctor, health officer	51 (10.6)
	Pharmacist, lab technician	100 (20.7)
	Nurse	311 (64.5)
	Other[a]	20 (4.1)
Education		
	BSc or Above	126 (26.1)
	Diploma	356 (73.9)
Training		
	Yes	364 (75.5)
	No	118 (24.5)

[a]Environmental health, dentistry, physiotherapy, and radiography

Computer Knowledge

Only 18.7% (90/482) of the respondents demonstrated good knowledge of computers in this study. Of them, few health professionals working at primary health centers (4/90, 4.4%) showed good computer knowledge compared to those working at government (21/90, 23.3%) and private (24/90, 26.7%) hospitals. The results are displayed in Figure 1.

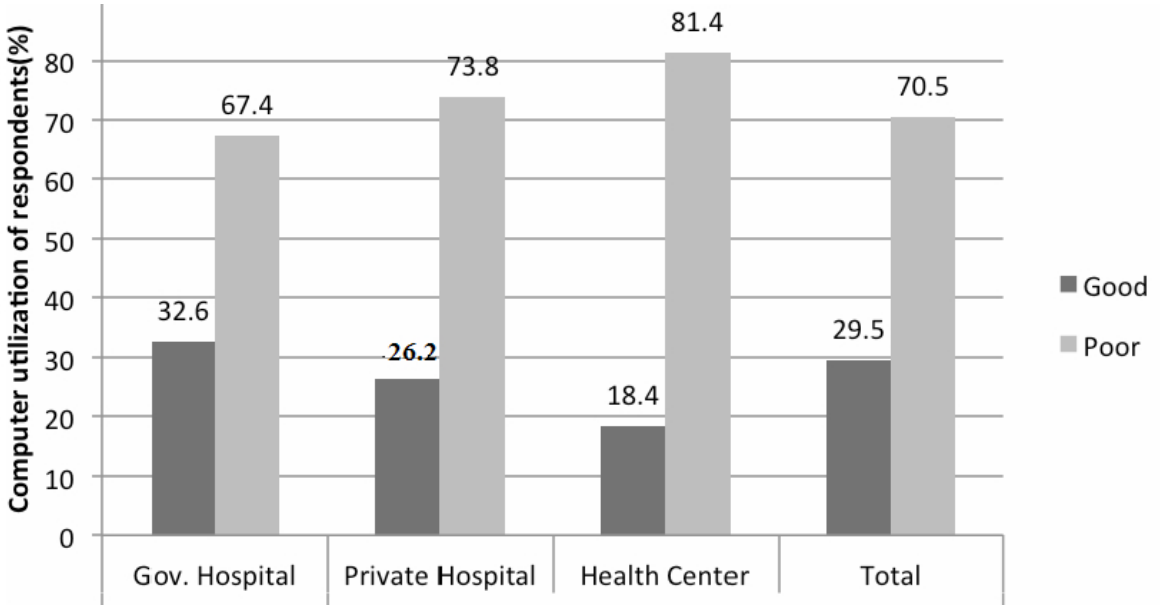
Figure 1. Knowledge of computers among health professionals.



Computer Utilization

A total of 29.5% of the respondents (142/482) had good utilization of computers. Participants working at government hospitals showed (115/353, 32.6%) good computer utilization, which was higher than those at private hospitals (11/42, 26.2%) and much higher than those at primary health centers (16/87, 18.4%). The results are shown in Figure 2 with more detail.

Figure 2. Utilization of computers among health professionals.



Factors Associated With Computer Utilization

With the multivariate logistic regression analysis done on computer utilization as dependent with other hypothesized independent variables, age, field of study, level of education, and computer training were found to be significantly associated with the computer utilization habits of health professionals.

To quantify each relationship, respondents who were younger (age 25-35) were approximately 3 times more likely to use computers than respondents aged 36 years and older (adjusted odds ratio [AOR]=3.06, 95% CI 0.57-5.37). Additionally, respondents who had previous computer training were 3.65 times more likely to use computers than those who did not have any kind of computer training (AOR=3.65, 95% CI 1.62-8.21). In the professional category, medical laboratory technicians and pharmacists were more likely to use computers than nurses (AOR=3.08, 95% CI 1.65-5.73). Additionally, those with higher levels of education were 2.78 times more likely to use computers than those with lower levels of education (AOR=2.78, 95% CI 1.43-5.40). The results are shown in Table 2.

Table 2. Factors associated with computer utilization among health professionals.

Predictor variables		Utilization		COR[a] (95% CI)	AOR (95% CI)
		Good	Poor		
Age					
	≤25	54	120	2.36 (1.31-4.25)	1.16 (0.40-3.34)
	26-35	46	72	5.36 (1.81-6.21)	3.06 (0.57-5.37)[c]
	≥36	19	100	1	1
Sex					
	Male	94	157	2.28 (1.51-3.43)	1.05 (1.01-2.69)
	Female	48	183	1	1
Marital status					
	Never married	75	127	1.87 (1.26-2.79)	1.40 (0.79-2.69)
	Married	67	213	1	1
Profession					
	Medical doctor, health officer	28	23	4.51 (2.44-8.35)	1.89 (0.77-4.61)
	Pharmacist, lab technician	39	61	2.37 (1.46-3.85)	3.08 (1.65-5.73)[c]
	Nurse	66	245	1	1
	Other[b]	9	11	3.03 (1.26-7.63)	1.30 (0.70-7.50)
Education					
	BSc or above	66	60	4.05 (2.63-6.24)	2.78 (1.43-5.40)[c]
	Diploma	76	280	1	1
Position					
	Institution head	6	4	3.93 (1.08-14.20)	1.97 (0.34-11.24)
	Team leader	28	53	1.38 (0.83-2.30)	1.70 (0.85-3.38)
	Care provider	108	283	1	1
Training					
	Yes	133	231	6.97 (3.42-14.21)	3.65 (1.62-8.21)[c]
	No	9	109	1	1
Service year					
	6-10	32	49	3.02 (3.27-24.77)	1.51 (1.92-22.02)
	11-15	9	45	2.76 (0.86-8.76)	1.28 (0.62-8.40)
	≥16	5	69	1	1

[a]Crude odds ratio

[b]Environmental health, dentistry, physiotherapy, and radiography

[c]Significant at $P<.05$.

Discussion

Principal Findings

The findings of this study show that computer knowledge and utilization was generally low and was lower for public health professionals who work in the primary health care centers. The results are lower compared to findings in previous studies [2,20,21], which might be attributed to a difference in study participants between the studies; previous studies only included health professionals working in hospitals while this study also includes health professionals working at the health centers, which have less computer access and information communication technology infrastructure.

The analysis of the determinant factors of computer utilization shows that age, field of study, level of education, and computer training were found to be significantly associated with computer

utilization. Among the factors, stronger association was found with computer training. The result is consistent with previous studies [20-23]. This result implies that providing trainings—not only about the specific eHealth software application which is going to be implemented but also generally about computers—can make a significance difference in system adaptation in health care.

In this study, we found that younger health professionals are more likely to use eHealth systems than older health professionals, which is consistent with other studies [24-27]. This implies that older employees need more assistance to adapt to and use the system.

Additionally, health professionals with advanced levels of education showed significantly better computer utilization than middle-level health professionals. The result is not surprising given the increasing number of computer-based tasks associated with further studies. Finding of this study was inconsistent with studies in India which showed that level of education was not significantly associated with computer utilization [21,22]. This may be due to training differences for health professionals in India and Ethiopia in which the basic computer courses in Ethiopia are more theoretical with few hours of practical lessons.

As skill is a main factor in eHealth success [28], interventions are needed to increase health professionals' knowledge and utilization. The Ministry of Health should provide training to the health professionals so that their knowledge can increase and their anxiety about technology can decrease. In addition, it is necessary to increase accessibility to computers, especially in primary care health centers, so health professionals can practice using computers in different activities before the main eHealth system is implemented.

In this study, knowledge and utilization habit measurements were self-reported, which might have some response bias. A further study complemented by qualitative approach is recommended to give more insight on how actual computer knowledge and utilization habits contribute to a better adoption of eHealth systems.

Limitations of the Study

This study did not address the attitude of health workers towards computers, which can influence their computer knowledge and utilization. Additionally, the information collected was self-perceived, which might have reporting bias. Future studies including attitude and actual practical use assessment are recommended. Additionally, the relationship between computer knowledge and use on eHealth success needs further investigation.

Conclusions

Computer knowledge and utilization habits of health professionals, especially those who work in primary health centers, were found to be low. Providing trainings and continuous follow-up are necessary measures to increase the likelihood of the success of implemented eHealth systems in those settings.

Acknowledgments

We are grateful for the data collectors and participants in this study. We would like to also thank Gondar University for funding the travel costs associated with this research.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Questionaire.

[PDF File (Adobe PDF File), 122KB - humanfactors_v2i1e4_app1.pdf]

References

1. Soar J, Gow J, Caniogo V. Sustainability of health information systems in developing countries: the case of Fiji. HIM J 2012;41(3):13-19. [Medline: 23087079]
2. Sukums F, Mensah N, Mpembeni R, Kaltschmidt J, Haefeli WE, Blank A. Health workers' knowledge of and attitudes towards computer applications in rural African health facilities. Glob Health Action 2014;7:24534 [FREE Full text] [Medline: 25361721]
3. Joan Dzenowagis, in collaboration with Gael Kernen. Connecting for health : global vision, local insight. In: Dzenowagis, Joan. , editor. Report for the World Summit on the Information Society. Geneva: World Health Organization; 2005.
4. Kimura K, Omole DW, Williams M. In: Chuhan-Pole P, Angwafo M, editors. Yes, Africa Can: Success Stories from a Dynamic Continent. Washington, DC: World Bank Publications; Jun 2011.
5. Ruxwana NL, Herselman ME, Conradie DP. ICT applications as e-health solutions in rural healthcare in the Eastern Cape Province of South Africa. HIM J 2010;39(1):17-26. [Medline: 20335646]
6. Simba DO. Application of ICT in strengthening health information systems in developing countries in the wake of globalisation. Afr Health Sci 2004 Dec;4(3):194-198.

7. Vital web consulting. Health Information system in developing countries. Health Information system in developing countries 2009:69 [FREE Full text]
8. Fraser HS, Blaya J. Implementing medical information systems in developing countries, what works and what doesn't. AMIA Annu Symp Proc 2010;2010:232-236 [FREE Full text] [Medline: 21346975]
9. Luna D, Otero C, Marcelo A. Health Informatics in Developing Countries: Systematic Review of Reviews. Contribution of the IMIA Working Group Health Informatics for Development. Yearb Med Inform 2013;8(1):28-33. [Medline: 23974545]
10. Azubuike MC, Ehiri JE. Health information systems in developing countries: benefits, problems, and prospects. J R Soc Promot Health 1999 Sep;119(3):180-184. [Medline: 10518358]
11. Ministry of Health. Health Sector Development Program IV of Ethiopia. In: Health Sector Development Program IV of Ethiopia. Addis Ababa: Ministry of Health; 2011:2010.
12. Binyam Tilahun,Atinkut Zeleke, Mengistu Kifle, Fleur Fritz. The Ethiopian national eHealth strategy and its alignment with the health informatics curriculum. Journal of Health Informatics in Africa 2015;2(2):47.
13. World Health Organization and International Telecommunication Union. National eHealth Strategy Toolkit. Publications of the World Health Organization and International Telecommunication Union 2012:5 [FREE Full text]
14. Tilahun B, Zeleke A, Fritz F, Zegeye D. New bachelors degree program in health informatics in Ethiopia: curriculum content and development approaches. Stud Health Technol Inform 2014;205:798-802. [Medline: 25160297]
15. Biruk S, Yilma T, Andualem M, Tilahun B. Health professionals readiness to implement electronic medical record system at three hospitals in Ethiopia: a cross sectional study. BMC Med Inform Decis Mak 2014 Dec 12;14(1):115 [FREE Full text] [doi: 10.1186/s12911-014-0115-5] [Medline: 25495757]
16. Vital wave Consulting. Health Information Systems in Developing Countries. vital web consulting Research Paper & Strategic Briefing 2009:69 [FREE Full text]
17. Fraser HS, Biondich P, Moodley D, Choi S, Mamlin BW, Szolovits P. Implementing electronic medical record systems in developing countries. Inform Prim Care 2005;13(2):83-95. [Medline: 15992493]
18. Asangansi OO, Adejoro OF, Makinde O. Computer use among doctors in Africaâ€¯: Survey of trainees in a Nigerian teaching hospital. Health informatics Dev Ctries., vol. 2, no. 1, pp. 10– 2008;2(1):10-14.
19. Cork RD, Detmer WM, Friedman CP. Development and initial validation of an instrument to measure physicians' use of, knowledge about, and attitudes toward computers. J Am Med Inform Assoc 1998;5(2):164-176 [FREE Full text] [Medline: 9524349]
20. Bello IS, Arogundade FA, Sanusi AA, Ezeoma IT, Abioye-Kuteyi EA, Akinsola A. Knowledge and utilization of Information Technology among health care professionals and students in Ile-Ife, Nigeria: a case study of a university teaching hospital. J Med Internet Res 2004 Dec 17;6(4):e45 [FREE Full text] [doi: 10.2196/jmir.6.4.e45] [Medline: 15631969]
21. Mohammed E, Andargie G, Meseret S, Girma E. Knowledge and utilization of computer among health workers in Addis Ababa hospitals, Ethiopia: computer literacy in the health sector. BMC Res Notes 2013;6:106 [FREE Full text] [doi: 10.1186/1756-0500-6-106] [Medline: 23514191]
22. Gour N, Srivastava D. Knowledge of computer among healthcare professionals of India: a key toward e-health. Telemed J E Health 2010 Nov;16(9):957-962. [doi: 10.1089/tmj.2010.0049] [Medline: 21091287]
23. Woreta SA, Kebede Y, Zegeye DT. Knowledge and utilization of information communication technology (ICT) among health science students at the University of Gondar, North Western Ethiopia. BMC Med Inform Decis Mak 2013;13:31 [FREE Full text] [doi: 10.1186/1472-6947-13-31] [Medline: 23452346]
24. Kipturgo MK, Kivuti-Bitok LW, Karani AK, Muiva MM. Attitudes of nursing staff towards computerisation: a case of two hospitals in Nairobi, Kenya. BMC Med Inform Decis Mak 2014;14:35 [FREE Full text] [doi: 10.1186/1472-6947-14-35] [Medline: 24774008]
25. Schumacher P, Morahan-Martin J. Gender, Internet and computer attitudes and experiences. Comput Human Behav 2001 Jan;17(1):95-110. [doi: 10.1016/S0747-5632(00)00032-7]
26. Palm JM, Colombet I, Sicotte C, Degoulet P. Determinants of user satisfaction with a Clinical Information System. AMIA Annu Symp Proc 2006:614-618 [FREE Full text] [Medline: 17238414]
27. Alison Gillwald, Anne Milek & Christoph Stork. Gender Assessment of ICT Access and Usage in Africa. Towards Evidence-based ICT Policy and Regulation 2010;1(5) [FREE Full text]
28. Dias C, Escoval A. Narrowing the Skills Gap for Innovation: An Empirical Study in the Hospital Sector. JMIR Hum Factors 2014 Sep 23;1(1). [doi: 10.2196/humanfactors.3598]

Abbreviations

COR: crude odds ratio
AOR: adjusted odds ratio

Edited by G Eysenbach; submitted 26.12.14; peer-reviewed by A Escoval, C Basch; comments to author 20.01.15; revised version received 25.01.15; accepted 03.02.15; published 26.03.15.

Please cite as:
Alwan K, Awoke T, Tilahun B
Knowledge and Utilization of Computers Among Health Professionals in a Developing Country: A Cross-Sectional Study
JMIR Human Factors 2015;2(1):e4
URL: http://humanfactors.jmir.org/2015/1/e4/
doi:10.2196/humanfactors.4184
PMID:27025996

Review

Using Eye Trackers for Usability Evaluation of Health Information Technology: A Systematic Literature Review

Onur Asan[1], PhD; Yushi Yang[2], MS

[1]Center for Patient Care and Outcomes Research, Department of Medicine, Medical College of Wisconsin, Milwaukee, WI, United States

[2]Department of Industrial Engineering, Clemson University, Clemson, SC, United States

Corresponding Author:
Onur Asan, PhD
Center for Patient Care and Outcomes Research
Department of Medicine
Medical College of Wisconsin
8701 Watertown Plank Road
Milwaukee, WI,
United States
Phone: 1 4149558847
Fax: 1 4149556689
Email: oasan@mcw.edu

Abstract

Background: Eye-tracking technology has been used to measure human cognitive processes and has the potential to improve the usability of health information technology (HIT). However, it is still unclear how the eye-tracking method can be integrated with other traditional usability methodologies to achieve its full potential.

Objective: The objective of this study was to report on HIT evaluation studies that have used eye-tracker technology, and to envision the potential use of eye-tracking technology in future research.

Methods: We used four reference databases to initially identify 5248 related papers, which resulted in only 9 articles that met our inclusion criteria.

Results: Eye-tracking technology was useful in finding usability problems in many ways, but is still in its infancy for HIT usability evaluation. Limited types of HITs have been evaluated by eye trackers, and there has been a lack of evaluation research in natural settings.

Conclusions: More research should be done in natural settings to discover the real contextual-based usability problems of clinical and mobile HITs using eye-tracking technology with more standardized methodologies and guidance.

(JMIR Human Factors 2015;2(1):e5) doi:10.2196/humanfactors.4062

KEYWORDS

health information technology; eye-tracking technology; usability evaluation

Introduction

Health information technology (HIT) systems are promising tools for improving quality, patient safety, and efficiency in health care systems [1-4]. This technology has been widely adopted due to governmental incentives, including funding, over the past few years [5]. However, despite powerful external forces driving the adoption of HIT, research has shown that physicians are still unsatisfied with, or resistant to, the technology [6] due to several unintended consequences from workflow and design-/usability-related problems. For example, one study reported that physicians felt that the standard reports produced by HIT systems actually reduced the usability and transparency of medical records [7]. To address usability issues and improve the design of HIT, usability evaluation research is necessary and becoming more prevalent [8,9]. Eye-tracking technology is one important tool that will be essential in such usability research.

Eye-tracking technology has been used to measure cognitive processes since the 1970s [10]. However, it has not been widely used for research purposes until recently, when the reduced cost of the equipment and user-friendly analysis tools made eye-tracking technology more readily available to researchers [11]. Eye-tracking technology is promising in HIT usability

research because of the close relation between visual stimuli and attentional mechanisms. Based on human information processing theory, people can only attend to a certain amount of visual stimuli at a time, due to a limited amount of mental resources [12]. Excessive information stimuli will result in mental overload that is correlated to physiological changes, such as pupil diameter [13]. Therefore, by tracking infrared light that is reflected by the human eye, we can understand a participant's mental load and cognitive state [14]. We can also detect the areas on a user interface that may capture users' attention and are processed by the human brain [15].

Two important measurements of eye-tracking technology are fixation and saccade [16]. Fixation has been operationally defined by previous researchers as a gaze that is longer than 300 milliseconds [17]. Fixation describes the moments when a human's eyes are relatively stationary, indicating the moments when the brain processes information received by the eyes [18]. Different patterns of fixation indicate different forms of human information processing. For example, high fixation rates usually indicate an area of great interest, which attracts the user's attention [19], whereas extremely long fixations indicate uncertainty and difficulties with information processing [10]. In addition, successive fixations are indications of inefficient visual search [20]. Saccades happen between fixations, when rapid eye movements shift attention from one target to another [18]. Saccade initiates when a critical cognitive event occurs and represents an attention shift [21].

Eye-tracking data can be integrated, synthesized, and visualized using software suites, such as commercially available analysis tools. Different types of visualizations, such as heat maps and gaze plots, communicate different types of information [22]. A heat map shows the observed areas and unobserved areas on an interface in different colors [23]. A gaze plot displays gaze motions by representing the sequence of saccades and fixations in the form of a scan path [24]. These visualizations are useful for explaining the user experience and usability of user interface design, and they help us make decisions on how to optimize the elements on that interface [25,26]. For example, heat maps and gaze plots have been used to determine certain areas of a webpage that attract the attention of viewers [27]. They have also been used to evaluate the usability of cartographic animations on interactive maps [28,29].

Based on the International Organization for Standardization (ISO) Standard, usability is the extent to which users can achieve a goal effectively, efficiently, and with satisfaction [30]. Due to the fact that eye-tracking measurement is closely related to attentional mechanisms and is able to accurately reveal cognitive processes, eye-tracking technology could play a more important role in this essential procedure for evaluating HIT. Yet, thus far it has been used minimally in usability evaluation studies. The objective of this literature review is to report and understand the current state of HIT usability evaluation studies that have used eye-tracker technology, and to envision the potential use of eye-tracking technology in future research.

Methods

Selection Strategy

We conducted a systematic online database search to identify articles published before September 2014 that were relevant to the aims of this study. Articles were included as indexed in four reference databases: Medline, Web of Science, ScienceDirect, and PsycINFO. Broad keyword searches were used to identify relevant articles in each database. Each initial search focused on one of three key components: (1) a word or phrase related to usability evaluation, (2) a word or phrase related to HIT, or (3) a word or phrase related eye-tracker technology.

Keywords related to usability evaluation included usability testing, user experience, user test, user-centered design, system design, interface design, and interaction design. Keywords related to HIT included health IT, health informatics, health technology, medical technology, eHealth, telemedicine, communication tools, educational technology, decision support technology, health app, and wearable technology. Keywords related to eye-tracker technology included eye-tracking technology, eye tracker, Tobii, Sensomotoric Instruments, eye movements, gaze, eye fixation, and saccade. We also identified potentially eligible articles by manual literature searches, by examining article reference lists and by searching in Google Scholar.

Inclusion and Exclusion Criteria

We initially defined the scope of the review by determining inclusion and exclusion criteria. Papers were included if they contained all of the following: (1) the research used eye-tracking technology as a data collection tool, (2) the research evaluated an HIT with users, and (3) the research explicitly mentioned the improvement of HIT usability based on the eye-tracking data. Papers were excluded if they (1) were not in English, (2) were published 10 or more years ago (ie, prior to 2004), (3) did not evaluate an HIT, (4) focused on technologies not related to health care, (5) used eye-tracker technology for some purpose other than data collection (ie, as an input device), or (6) did not mention any indications of the system usability based on the eye-tracker data.

Analysis

Based on the methods-description approach, we analyzed the selected papers that met the inclusion criteria [31]. Key article characteristics were recorded using a template with the following sections: title, author, purpose, and key findings [31]. After the creation of the table, we captured key data by coding as the recurrent topics. Coding is an analytical process that allows the articles to be categorized based on factors that are thought to be important [32]. Through the coding process, the following topics were explored: the research question answered by eye-tracking data, types of health IT to be evaluated, evaluation apparatus, eye-tracker measurement and analysis, and how eye-tracker technology is combined with other usability evaluation methods.

Results

Overview

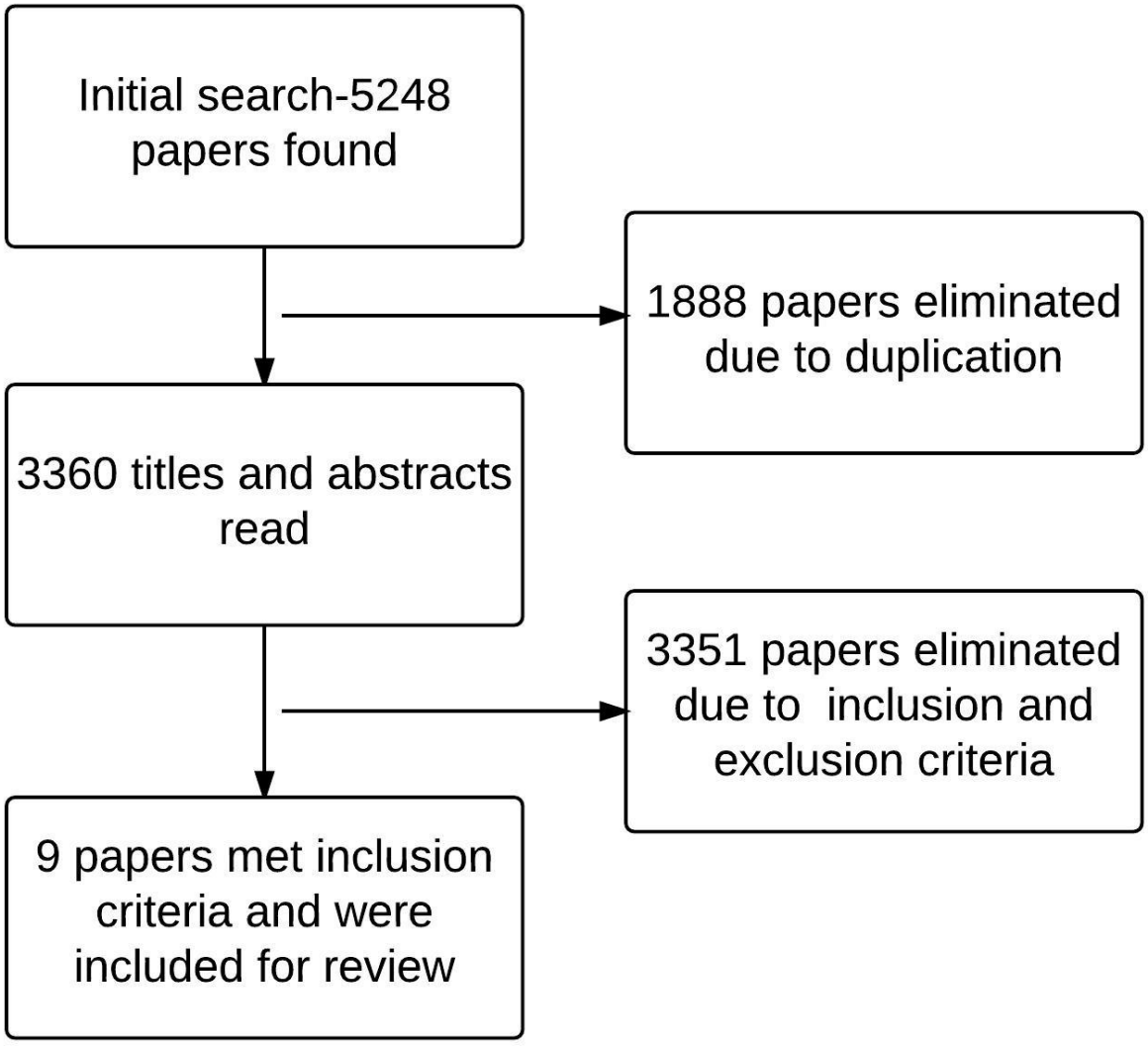
A total of 5248 papers were found by using the search terms and databases described above. Of these, 1888 papers were removed due to duplication. After reviewing the titles and abstracts based on inclusion and exclusion criteria, we eliminated another 3351 papers. This resulted in a total of 9 papers remaining for this review (see Figure 1). An overview of the 9 papers can be found in Table 1 [33-41]. It is important to note that 2 of the 9 papers are from the same project [35,36]. We included both because they fit the inclusion criteria. Of these, 1 paper describes one of the earliest studies using eye-tracking technology to evaluate the usability of a computer application [35], and the other is a complete report of the whole user-centered design process, which reflected more information on the entire research context [36].

All selected papers discussed user evaluation of a type of HIT using eye-tracking technology as a data collection tool. Of the 9 papers, 3 of them (33%) mainly discussed a usability evaluation of an HIT using eye-tracker technology [33,35,40]. Of the 9 papers, 3 of them (33%) presented an entire user-centered design process and discussed the usability evaluation of an HIT using eye-tracking technology as one part of the paper [34,36,41]. For instance, 1 study discussed how focus groups were used as a way to develop a quality-of-life support prototype, and then evaluated the usability of the prototype using eye-tracker technology [41]. The main purpose of the 3 remaining papers was not usability evaluation of the HIT, however, the eye-tracking data derived from the user evaluation clearly provided a basis for usability improvement [37-39]. For instance, 1 study reported that providers did not recognize patient-identity errors on a computerized provider order entry (CPOE), even if the eye-tracking data indicated that they looked at the area that contained errors [37]. These results have been translated to usability improvement recommendations for the system, for example, to make the important identity information more salient on the interface.

Table 1. Summaries of papers used in the review.

Author and reference	Title	Purpose	Key findings
Bansback et al [33]	Development and preliminary user testing of the DCIDA (Dynamic computer interactive decision application) for 'nudging' patients towards high quality decisions	To develop and test a computer application that enhances conventional patient decision aids so that common decision errors made by patients can be reduced.	The Dynamic Computer Interactive Decision Application (DCIDA) version of patient decision aids was understandable to users and it was able to help users focus on attributes that are of individual importance to them.
Barkana and Acik [34]	Improvement of design of a surgical interface using an eye tracking device.	To use eye-tracking technology to improve the design of a surgical interface to obtain the optimum configuration.	The interface of the early version of a surgical interface was redundant. With two larger scans at higher spatial resolution on the interface, participants were able to complete tasks more quickly, and the visual acquisition corresponded more to the natural visual search.
Eghdam et al [35]	Combining usability testing with eye-tracking technology: evaluation of a visualization support for antibiotic use in intensive care	To investigate if Infobiotika supports efficient and effective navigation and to observe the user's navigation paths, visual scan patterns, and distribution of visual attention.	Infobiotika was effective and efficient in terms of navigation support, and was a learnable product for intensive care unit (ICU) physicians.
Forsman et al [36]	Integrated information visualization to support decision making for use of antibiotics in intensive care: design and usability evaluation	To investigate the role of visualization as a method to support intensive care physicians' decision making about antibiotic use, analyze users' work processes and information needs, develop an interactive tool for integrated information visualization, and perform usability testing.	The visualization tool was usable for supporting ICU physicians in antibiotic use. Physicians had increased awareness of a patient's infection-related data and felt more in control of the situation.
Henneman et al [37]	Providers do not verify patient identity during computer order entry	To determine the frequency of verifying patient identity in an emergency department (ED) during computerized provider order entry (CPOE).	Medical providers did not usually verify patient identity prior to selecting the patient from the list and ordering tests. They often did not recognize patient-identity errors in the system.
Kules and Xie [38]	Older adults searching for health information in MedlinePlus – an exploratory study of faceted online search interfaces	To examine how searchers interact with a faceted Web-search interface.	Faceted interfaces played a substantial role in participants' use of the search result pages. The severity of the health condition affected the use of faceted interfaces.
Liu et al [39]	The use of illustration to improve older adults' comprehension of health-related information: Is it helpful?	To examine whether explanatory illustrations can improve older adults' comprehension of written health information.	Older adults had difficulties understanding the illustrations as well as integrating the illustrations with the text. Older adults did not benefit from the use of illustration.
Rashid et al [40]	Preliminary usability testing with eye tracking and FCAT analysis on occupational safety and health websites	To measure effectiveness, efficiency, and satisfaction of the Occupational Safety and Health (OSH) website, and to gather user feedback.	Eye-tracker data and user feedback helped identify usability problems of three OSH websites.
Wolpin et al [41]	Development and usability testing of a web-based cancer symptom and quality-of-life support intervention	To develop a user-centered prototype, and assess user preferences from usability testing of a revised prototype of the Electronic Self-Report Assessment for Cancer-II (ESRAC 2.0) project.	An application was developed that integrated the patients' needs through the methods of participatory design, usability testing, and iterative development.

Figure 1. Flow diagram of the study selection process.



What Research Questions Are Answered by Eye-Tracker Technology?

We identified different research questions that are answered by eye-tracker technology in the selected papers. The first question that can be answered by eye-tracker technology is whether the user experience and performance using an HIT has been improved based on the eye-gaze patterns, which primarily reflects the system effectiveness, efficiency, and user satisfaction [33-36,40,41]. The second question that can be answered by eye-tracker technology is how people use visual cues in the decision-making process, which primarily reflects the linkage between human visual stimulus and cognitive processing [37]. The third question that can be answered by eye-tracker technology is how information is processed differently under different circumstances, such as age and health conditions, which primarily reflects the variability of human performance [38,39].

What Types of Health Information Technologies Were Evaluated Using an Eye Tracker?

We identified different types of HITs in the selected papers. In terms of functionality, the technologies included online health information website interfaces [38-40], surgical interfaces [34], decision support systems [33,35,36], computerized provider order entry systems [37], and symptom and quality-of-life information systems [41]. In terms of target users, the health care information technologies were for the general public [38-40], patients [33,41], and physicians [34-37].

What Is the Experimental Apparatus of the Usability Test?

We identified different experimental apparatuses of the user tests. Researchers evaluated HITs in the forms of developed computer website/application [33,34,38,40], simulated prototype [35,36,41], and screenshots [37,39]. Researchers used three different kinds of eye trackers to collect data: on-screen eye trackers (Tobii T60 and T120) [33,38,40,41], mobile eye trackers that are external to a personal computer (Sensomotoric

Instruments [SMI] 500, Tobii X-60 and X120) [34-36], and head-mounted eye trackers [37,39]. Experiments were conducted either in a usability lab room or a meeting room. None of the experiments were conducted in the natural setting.

Out of the 9 papers, 2 of them (22%) reported a failure to collect eye-tracking data during the usability test [37,39]. Of those 2 papers, 1 of them reported that data for 12 out of 250 patient identification scenarios were not recorded due to failures in the eye-tracking system [37]. The other paper reported that the eye tracker was not able to perform for one-third of the older adult participants [39]. Both papers used a head-mounted eye tracker for data collection.

What Did the Eye Tracker Measure and How Was Data Analyzed?

We identified three basic eye-tracker measurements in our selected papers. The measurements included fixation duration [33,34,38-40], the locations of eye movement [35-37], and the fixation count in an area of interest [34,38]. Some papers included two measurements, focusing on both fixation duration and number of fixations in an area of interest [34,38]. Three basic methods were also used to analyze the eye-tracker measures in the selected papers, including heat map [33,34,40], gaze plot [35,36,41], and statistical analysis [34,37-39]. Generally, a heat map is used when fixation-duration data is collected [33,34,38], a gaze plot is used when the location of eye-movement data is collected [35,36], and statistical analysis is used when fixation-duration data is collected [34,38,39]. The heat map and gaze plot are qualitative methods for understanding the observed areas and gaze motions on an interface. Statistical analysis is a quantitative method to examine the effects of two different versions of a design or two different user groups on the task completion time.

How Is Eye-Tracker Technology Combined With Other Usability Methods?

The selected papers also showed other usability evaluation methods that are combined with eye-tracker data to explore usability problems in HIT systems. The methods include the System Usability Scale (SUS) [33,35,36], the think-aloud protocol [33,38,41], the National Aeronautics and Space Administration Task Load Index (NASA-TLX) and Short Post-Assessment Situational Awareness (SPASA) questionnaire [34], posttest interviews [36], metrics measurement [39], and Feedback Capture After Task (FCAT) [39]. There are two different types of think-aloud evaluations: concurrent think aloud, which encourages participants to tell what they think while using the program, and retrospective think aloud (RTA), which asks participants to verbalize their thoughts afterwards. Researchers in selected papers used concurrent think aloud [41], RTA [38], and a combination of both [33]. Table 2 shows the research questions that were answered by eye-tracker technology in the selected papers.

Table 2. Summary of research questions.

Questions and answers		Reference
Q1: What research questions are answered by eye-tracker technology?		
	System effectiveness, efficiency, and user satisfaction	[33-36,40,41]
	Linkage between human visual stimulus and cognitive processing	[37]
	The variability of human performance	[38,39]
Q2: What types of HITs were evaluated using an eye tracker?		
Technology type by functionality		
	Health information website interfaces	[38-40]
	Surgical interfaces	[34]
	Decision support systems	[33,35,36]
	Computerized provider order entry systems	[37]
	Symptom and quality-of-life information systems	[41]
Technology type by target users		
	General public	[38-40]
	Patients	[33,41]
	Physicians	[34-37]
Q3: What is the experimental apparatus of the usability test?		
Experimental apparatus by technology		
	Developed computer program	[33,34,38,40]
	Simulated prototype	[35,36,41]
	Screenshots	[37,39]
Experimental apparatus by eye tracker		
	On-screen eye tracker	[33,38,40,41]
	Mobile eye tracker	[34-36]
	Head-mounted eye tracker	[37,39]
Q4: What did the eye tracker measure and how was data analyzed?		
Eye-tracker data collected		
	Fixation duration	[33,34,38-40]
	Eye movement location	[35-37]
	Fixation count in area of interest	[34,38]
Eye-tracker data analyzed		
	Heat map	[33,34,40]
	Gaze plot	[35,36,41]
	Statistical analysis	[34,37-39]
Q5: How is eye-tracker technology combined with other usability methods?		
	Think-aloud protocol	[33,38,41]
	System Usability Scale	[33,35,36]
	Questionnaire	[34]
	Posttest interview	[36]
	Metrics measurement	[39]
	Feedback Capture After Task	[39]

Discussion

Principal Findings

The purpose of this literature review was to examine usability evaluations of any type of HIT using eye-tracking technology. This review also aimed to identify the research gap and potential uses of eye-tracker technology in future HIT research. This review was conducted based on the inclusion and exclusion criteria specified in the Methods section. Based on the results, we determined that, although eye trackers provide rich data for the improvement of HIT systems, the use of eye trackers for usability evaluation of HITs is still in its infancy, as only 9 papers were found that fit within the inclusion criteria.

We organized the results into five main questions: (1) What research questions are answered by eye-tracking technology?, (2) What types of health care information technologies were evaluated using an eye tracker?, (3) What was the experimental apparatus of usability evaluation?, (4) What did the eye trackers measure and how was data analyzed?, and (5) How was eye-tracker technology combined with other usability methods?

Papers that were included in this review had different purposes and research goals. The types of HITs evaluated were limited, resonating with our finding that the use of eye trackers for the evaluation of health IT is in an early stage. However, eye trackers are becoming a promising tool for usability studies, as demonstrated by the increasing number of research studies in recent years. We also found that researchers used various means of data collection and analysis using eye trackers. On the one hand, this demonstrates the rich variety of data that can be captured by eye trackers and the flexibility of interpretation of eye-tracker data. On the other hand, it shows the lack of a consensus on how to conduct user evaluation of HITs using eye trackers at this stage. In addition, we found that eye-tracking technology, as a part of usability evaluation methodology, was supplemented by other traditional methods. Generally, eye-tracking data can reveal the patterns of user difficulties when completing tasks using HIT, while other supplemental inquiries are used to unfold the reasons behind those patterns. Therefore, eye-tracking technology has to integrate with other techniques, as most physiology measurements do, because eye-tracking technology alone cannot tell the entire story.

Different Research Questions

The reviewed papers reflected different research questions that were answered by eye-tracking technology. Of the 9 papers, 6 of them (67%) were directly related to the system usability, focusing on the efficiency, effectiveness, and satisfaction when completing tasks with a specific HIT. Of these, 1 paper was related to the examination of a gap between visual and cognitive process. For example, a user missed information because he/she did not pay attention to it, even if eye-tracking technology suggested that the user had seen that information [37]. Another 2 papers (22%) out of 9 were related to the evaluation of different gaze patterns under different circumstances. For example, age had been identified as a factor for processing information [39]. Although the research questions were different, all of these studies commented on how eye-tracking data might have direct or indirect implications for the usability improvement of the evaluated HIT.

Limited Types of Health Information Technologies

The reviewed papers involved five different types of HITs, including 3 out of 9 papers (33%) evaluating health information website interfaces, 3 papers (33%) evaluating decision support systems, 1 paper (11%) evaluating a surgical interface for physicians, 1 paper (11%) evaluating a computerized provider order entry system for physicians, and 1 paper (11%) evaluating a symptom and quality-of-life information system for physicians. The reviewed papers involved three different types of users, including 3 papers (33%) for general public health IT, 2 papers (22%) for patients, and 4 papers (44%) for physicians. Thus far, eye trackers have been used most often to evaluate health information website interfaces. This indicates that evaluating a website interface using eye-tracker analysis may provide rich theoretical guidance and reveal available practices that researchers can refer to [42,43]. Moreover, the methods for evaluating a website interface are familiar to usability specialists.

However, there is much potential for eye-tracker technology to be applied to other types of health IT as well. One particular aspect of health IT that lacks usability research using eye-tracker technology is electronic health record (EHR) systems. EHR systems have helped to revitalize physician and nursing practice, and have the potential for positive impact on clinical processes in terms of efficiency, productivity, and patient safety [44]. Health care providers' attitudes toward EHR systems have been assessed and results showed that a majority recognized the positive influence of EHR systems in terms of decreased workload, improved quality of documentation, and electronic charting [45]. However, some other studies also reported a negative impact of EHRs, such as workflow interruptions and introduction of new errors because of usability factors, which have also been identified as a major barrier for successful EHR implementation [46,47]. Eye-tracking technology can also be used to identify usability problems to improve the design in a better way.

Another gap exists in the application of eye-tracking technology to usability studies of novel consumer HITs. Health apps and devices are becoming prevalent in the market. Devices such as the iPad, iPhone, iPod Touch, and Apple Watch have been the target devices to provide a richer and more convenient user experience of health care information technology [48]. Wearable interfaces and Web-based activity-monitoring systems are popular in the current market for the encouragement, persuasion, and guidance of healthy lifestyles. Because of the smaller size of these screens, there are increased difficulties for users to operate these systems and for designers to maximize the available screen area effectively [49]. Also, users expect to interact with these HITs in ways that are consistent with other technologies, without the need to read instructions. Eye-tracking technology has the ability to examine whether this has been achieved [50]. In that regard, eye-tracking data would be very helpful in understanding how users interact with those technologies and in providing designers with the basis to make improvements.

Lack of Research in a Natural Setting

In terms of the prototypes that were evaluated in the reviewed papers, a majority of them (4/9, 44%) evaluated developed computer programs that have already been adopted in health care systems. Of these papers, 2 of them evaluated screenshots of the real websites, and 2 of them evaluated usability using a simulated prototype in high fidelity. Unlike many other usability techniques, such as the formative usability evaluation approach, that are primarily introduced in the early phase of the user-centered design process, we found almost all the papers evaluated HITs at a very late phase or even after implementation, as a summative approach. The benefit of doing a summative usability evaluation is that researchers are able to create an approximation of the actual use scenario of HITs. Compared with low-fidelity, nonfunctional prototypes, such late-phase testing is more likely to uncover real usability problems [51]. However, even such high-fidelity approximations fall short of researching HIT use in the natural setting. The health care system is a sociotechnical system with a complex structure, complex dynamics, and multiple stakeholders [52]. Not until health care providers work in the real environment can many organizational issues emerge, such as patient privacy, workflow complexity, and disruptions [52-57]. Those factors influence the usability of HIT in ways that cannot be captured by lab-based evaluations. Therefore, an ecological gap is a particular concern for HIT evaluation representing the differences of user study results between the lab and the real setting [58]. Because of this, we believe there are certain usability problems of HIT that can only be discovered in the field within the real context of HIT use.

Unfortunately, at this point there has been no usability evaluation conducted using eye-tracking technology in real settings. All of the reviewed papers conducted user studies in a meeting room or a usability lab. Possible reasons for this gap could be the mobility limitations of eye-tracker technology, the possible intrusion of such technology on work, technical difficulties, and the calibration process of the eye-tracking equipment. Nearly half of the papers (4/9, 44%) used eye trackers that were embedded within a computer screen, which are impossible to move into real settings. Of the 9 papers, 2 of them (22%) used head-mounted eye trackers, which are easy to move but intrusive to the health care provider's work if the evaluation is conducted in the field, negatively influencing the work in that time-sensitive environment. Moreover, such head-mounted trackers are more likely to have technological difficulties, which risk accurate data collection. Of the 9 papers, 3 of them (33%) used mobile eye trackers, which are probably the best equipment to be incorporated into field research in the real HIT setting. However, the calibration process may add additional steps to the already complex workload of nurses or physicians. Moreover, it is unlikely that a health care provider will stay in one place for a long period of time, and their movements will disrupt the calibration [59]. Despite this, we still believe in the necessity and value of conducting real-life usability evaluations of health IT using eye trackers. We expect advancements in eye-tracking technology to address this obstacle. For example, a new technology—Glasses—is capable of collecting data in real settings without the problems of calibration or too much intrusion on the health care provider's work.

Gaps of Eye-Tracker Data Analysis

We found that the eye-tracker measurements in the reviewed papers were mainly fixation and saccade, which supports the finding by Poole and Ball [18]. More than half of the reviewed papers (5/9, 56%) collected fixation-duration data. For example, researchers used fixation duration as an indicator of the efficiency of human interaction with the surgical interface [34]. Of the 9 papers, 3 of them (33%) collected eye-movement locations. Of these, 2 papers collected both the fixation duration and fixation count in areas of interest. For example, researchers evaluated a Dynamic Computer Interactive Decision Application, using fixation number and fixation duration as indicators of attributes on the DCIDA [33]. We found that certain quantitative eye-tracker data are more favored by researchers, such as fixation duration and fixation count. Qualitative data collection and analysis appeared less frequently in the reviewed papers, which corresponds to the finding by Yen and Bakken [9].

Qualitative analysis is becoming prevalent partly because of the improvement of software suites, making the analysis easier and less intensive. Of the 9 papers, 3 of them (33%) translated the data into qualitative visualization, such as heat maps and gaze plots. While statistical analysis is powerful in comparing completion time and errors, it is only part of the usability evaluation. For a full usability evaluation, we believe the qualitative data in visualization can illustrate more usability problems. Using a heat map, it is easy to determine if specific content is usable or not. Using a gaze plot, it is possible to determine if users follow an efficient and predetermined route when searching for specific information on the interface.

However, we found that the interpretation of such visualizations lacked scientific guidance based on an established theoretical method, so interpretations tended to seem arbitrary and subjective. At this point, researchers are struggling to find a theory or a commonly used procedure to guide the interpretation of heat maps and gaze plots. Therefore, we expect that in the future a more structured system of interpretation will be developed for heat maps and gaze plots.

Opportunities for Integration

With the visualization of eye-tracker data, researchers can identify the areas of an interface that have created difficulties in participants' minds. However, based solely on the eye-tracker data, there is no way to understand the precise cognitive reasons behind a participant's eye-gaze patterns. For example, there might be many possibilities for an eye fixation, such as fatigue, distraction, confusion, and engagement [18,60]. Therefore, researchers will have to integrate other quantitative and qualitative research methods with eye-tracking research in order to understand why people behave in a particular way. Of the 9 papers, 7 of them (78%) used other methods along with the eye tracker, some using more than one method. Of these, 3 papers used the think-aloud method, 3 papers used the SUS, 1 paper used the NASA-TLX and the SPASA questionnaire, 1 paper

used a posttest interview, 1 paper used metrics measurements, and 1 paper used FCAT.

It is interesting that there are three different think-aloud methods used in the reviewed papers: concurrent think aloud, RTA, and a hybrid of both. The concurrent think-aloud method is the traditional method widely accepted and applied by usability evaluation researchers. It is a method that asks participants to verbalize their thoughts while interacting with the system [61]. However, the method has received criticism because the verbal process requires attention and may distract the participants [62]. Additionally, during the think-aloud method, users usually have the temptation to look at the researcher for conversation, which has the risk of disrupting the calibration of eye-tracking technology, thus causing researchers to lose eye-tracking data [63]. RTA records participants' eye movements during the usability test session and then asks them to verbalize their thoughts afterward while watching the gaze-plot animation [64]. Research has shown that RTA enhances the validity and reliability of the usability evaluation results [65]. However, RTA does have some identified limitations, due to the limited capability of eye-tracker technology, which we need to be aware of. For example, eye trackers are not able to capture peripheral vision data. Although our peripheral vision is in low resolution, that still accounts for part of our visual input [11]. Similarly, orphan fixation can happen when the user is making some unintentional fixation or when the user looks at an area, but attention is somewhere else [59]. When researchers present this to the participant, it can surprise the participants and, therefore, distract them from an efficient RTA process [59].

Another reviewed paper used a hybrid of concurrent think-aloud and RTA methods [33]. The researchers asked the participants to think aloud while completing the task. However, if they could not think aloud about a particular page within 10 seconds, they would be asked to reflect after the task session. This method is superior because the participants have the opportunity to verbalize immediate thoughts during the evaluation session, but also have the opportunity to review and think more deeply after the test.

Conclusions

Although eye tracking is a promising technology, the application of eye-tracking technology to health IT usability evaluation is still in its infancy, with limited theoretical guidance and practice. Therefore, we reviewed papers that were related to usability evaluations of HIT using an eye tracker, to understand the current state, identify the gaps, and envision future research. There is no doubt that eye-tracker technology would be able to provide valuable data if well-integrated with other traditional usability evaluation methodologies. However, the lack of field research of clinical and mobile HITs in natural settings is a huge gap that needs to be filled. Scientific guidance is also needed for the interpretation of eye-tracking visualizations. Eye trackers can play a significant role in the future of usability evaluations of HIT if they are used effectively and correctly.

Acknowledgments

This research was supported by the Center for Patient Care and Outcomes Research at the Medical College of Wisconsin. We would also like to thank Beth Schewe for her contribution to the paper.

Conflicts of Interest

None declared.

References

1. Bates DW, Gawande AA. Improving safety with information technology. N Engl J Med 2003 Jun 19;348(25):2526-2534. [doi: 10.1056/NEJMsa020847] [Medline: 12815139]
2. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. PLoS Med 2011 Jan;8(1):e1000387 [FREE Full text] [doi: 10.1371/journal.pmed.1000387] [Medline: 21267058]
3. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Ann Intern Med 2006 May 16;144(10):742-752. [Medline: 16702590]
4. 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]
5. Jha AK, Doolan D, Grandt D, Scott T, Bates DW. The use of health information technology in seven nations. Int J Med Inform 2008 Dec;77(12):848-854. [doi: 10.1016/j.ijmedinf.2008.06.007] [Medline: 18657471]
6. Ford EW, Menachemi N, Peterson LT, Huerta TR. Resistance is futile: but it is slowing the pace of EHR adoption nonetheless. J Am Med Inform Assoc 2009 May;16(3):274-281 [FREE Full text] [doi: 10.1197/jamia.M3042] [Medline: 19261931]
7. Ash JS, Berg M, Coiera E. Some unintended consequences of information technology in health care: the nature of patient care information system-related errors. J Am Med Inform Assoc 2004 Apr;11(2):104-112 [FREE Full text] [doi: 10.1197/jamia.M1471] [Medline: 14633936]
8. Rahimi B, Vimarlund V. Methods to evaluate health information systems in healthcare settings: a literature review. J Med Syst 2007 Oct;31(5):397-432. [Medline: 17918694]

9. Yen PY, Bakken S. Review of health information technology usability study methodologies. J Am Med Inform Assoc 2012 Jun;19(3):413-422 [FREE Full text] [doi: 10.1136/amiajnl-2010-000020] [Medline: 21828224]
10. Jacob RJK, Karn KS. Eye tracking in human-computer interaction and usability research: ready to deliver the promises. In: Radach R, Hyona J, Deubel H, editors. The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research. Amsterdam, The Netherlands: Elsevier Science BV; 2003:573-603.
11. Schall A, Bergstrom J. Introduction to eye tracking. In: Bergstorm J, Schall A, editors. Eye Tracking in User Experience Design. Waltham, MA: Morgan Kaufmann; 2014:3-26.
12. Wickens CD, Hollands JG, Parasuraman R, Banbury S. Engineering Psychology and Human Performance. 4th edition. Boston, MA: Pearson; 2012.
13. Poock GK. Information processing vs pupil diameter. Percept Mot Skills 1973 Dec;37(3):1000-1002. [doi: 10.2466/pms.1973.37.3.1000] [Medline: 4764491]
14. Palinko O, Kun AL, Shyrokov A, Heeman P. Estimating cognitive load using remote eye tracking in a driving simulator. In: Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications. 2010 Presented at: ETRA '10; March 22-24, 2010; Austin, TX p. 141-144. [doi: 10.1145/1743666.1743701]
15. Richardson DC, Spivey MJ. Eye tracking: Research areas and applications. In: Wnek GE, Bowlin GL, editors. Encyclopedia of Biomaterials and Biomedical Engineering. London, UK: Taylor & Francis; 2004:573-582.
16. Bruneau D, Sasse MA, McCarthy J. The eyes never lie: the use of eyetracking data in HCI research. In: Proceedings of the CHI2002 Workshop on Physiological Computing. 2002 Presented at: CHI2002 Workshop on Physiological Computing; April 21, 2002; Minneapolis, MN URL: http://hornbeam.cs.ucl.ac.uk/hcs/people/documents/Angela%20Publications/2002/CHI_Physio.computing_Final%20%281%29_revised.pdf
17. Djamasbi S, Tullis T, Hsu J, Mazuera E, Osberg K, Bosch J. Gender preferences in Web design: usability testing through eye tracking. In: Proceedings of the Thirteenth Americas Conference on Information Systems (AMCIS). 2007 Presented at: Thirteenth Americas Conference on Information Systems (AMCIS); August 9-12, 2007; Keystone, CO URL: http://digitalcommons.wpi.edu/cgi/viewcontent.cgi?article=1022&context=uxdmrl-pubs
18. Poole A, Ball L. Eye tracking in human-computer interaction and usability research: Current status and future. In: Ghaoui C, editor. Encyclopedia of Human-Computer Interaction. Hershey, PA: Idea Group Reference; 2005:211-219.
19. Velazquez CE, Pasch KE. Attention to food and beverage advertisements as measured by eye-tracking technology and the food preferences and choices of youth. J Acad Nutr Diet 2014 Apr;114(4):578-582. [doi: 10.1016/j.jand.2013.09.030] [Medline: 24332473]
20. Goldberg JH, Kotval XP. Computer interface evaluation using eye movements: methods and constructs. Int J Ind Ergon 1999 Oct;24(6):631-645. [doi: 10.1016/S0169-8141(98)00068-7]
21. McConkie GW, Yang SN. How cognition affects eye movements during reading. In: Radach R, Hyona J, Deubel H, editors. The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research. Amsterdam, The Netherlands: Elsevier Science BV; 2003:413-427.
22. Nielsen J, Pernice K. Eyetracking Web Usability. Berkeley, CA: New Riders; 2009.
23. Špakov O, Miniotas D. Visualization of eye gaze data using heat maps. Electronics and Electrical Engineering 2007:55-58 [FREE Full text]
24. Shrestha S, Lenz K. Usability News. Wichita, KS: Software Usability Research Laboratory (SURL), Wichita State University; 2007. Eye gaze patterns while searching vs. browsing a website URL: http://psychology.wichita.edu/surl/usabilitynews/91/eyegaze.asp [accessed 2014-11-19] [WebCite Cache ID 6UCzSc958]
25. Reingold EM. Eye tracking research and technology: towards objective measurement of data quality. Vis cogn 2014 Mar;22(3):635-652 [FREE Full text] [doi: 10.1080/13506285.2013.876481] [Medline: 24771998]
26. Manhartsberger M, Zellhofer N. Eye tracking in usability research: What users really see. In: Empowering Software Quality: How Can Usability Engineering Reach These Goals?. 2005 Presented at: Usability Symposium; November 8, 2005; Vienna, Austria p. 141-152 URL: http://www.usability.at/ueberuns/images/EyetrackinginUsability.pdf
27. Djamasbi S, Siegel M, Tullis T. Generation Y, web design, and eye tracking. Int J Hum Comput Stud 2010 May;68(5):307-323. [doi: 10.1016/j.ijhcs.2009.12.006]
28. Dong W, Liao H, Xu F, Liu Z, Zhang S. Using eye tracking to evaluate the usability of animated maps. Sci China Earth Sci 2013 Oct 10;57(3):512-522. [doi: 10.1007/s11430-013-4685-3]
29. Opach T, Nossum A. Evaluating the usability of cartographic animations with eye-movement analysis. In: Proceedings of the 25th International Cartographic Conference. 2011 Presented at: 25th International Cartographic Conference; July 3-8, 2011; Paris, France URL: http://icaci.org/files/documents/ICC_proceedings/ICC2011/Oral%20Presentations%20PDF/A3-Visualisation%20efficiency/CO-005.pdf
30. Abran A, Khelifi A, Witold S, Seffah A. Usability meanings and interpretations in ISO standards. Software Quality Journal 2003;11(4):325-338. [doi: 10.1023/A:1025869312943]
31. Cooper HM. Synthesizing Research: A Guide for Literature Reviews. 3rd edition. Thousand Oaks, CA: SAGE Publications; 1998.
32. Berg BL, Lune H. Qualitative Research Methods for the Social Sciences. 8th edition. New York, NY: Pearson; 2011.

33. Bansback N, Li LC, Lynd L, Bryan S. Development and preliminary user testing of the DCIDA (Dynamic computer interactive decision application) for 'nudging' patients towards high quality decisions. BMC Med Inform Decis Mak 2014;14:62 [FREE Full text] [doi: 10.1186/1472-6947-14-62] [Medline: 25084808]
34. Erol Barkana D, AçÄ±k A. Improvement of design of a surgical interface using an eye tracking device. Theor Biol Med Model 2014 May 7;11 Suppl 1:S4 [FREE Full text] [doi: 10.1186/1742-4682-11-S1-S4] [Medline: 25080176]
35. Eghdam A, Forsman J, Falkenhav M, Lind M, Koch S. Combining usability testing with eye-tracking technology: evaluation of a visualization support for antibiotic use in intensive care. Stud Health Technol Inform 2011;169:945-949. [Medline: 21893885]
36. Forsman J, Anani N, Eghdam A, Falkenhav M, Koch S. Integrated information visualization to support decision making for use of antibiotics in intensive care: design and usability evaluation. Inform Health Soc Care 2013 Dec;38(4):330-353. [doi: 10.3109/17538157.2013.812649] [Medline: 23957739]
37. Henneman PL, Fisher DL, Henneman EA, Pham TA, Mei YY, Talati R, et al. Providers do not verify patient identity during computer order entry. Acad Emerg Med 2008 Jul;15(7):641-648. [Medline: 19086323]
38. Kules B, Xie B. Older adults searching for health information in MedlinePlus â€“ An exploratory study of faceted online search interfaces. In: Proceedings of the 74th ASIS&T Annual Meeting. 2011 Presented at: ASIS&T 2011; October 9-12, 2011; New Orleans, LA p. 9-13 URL: http://www.asis.org/asist2011/proceedings/submissions/137_FINAL_SUBMISSION.pdf
39. Liu CJ, Kemper S, McDowd J. The use of illustration to improve older adults' comprehension of health-related information: is it helpful? Patient Educ Couns 2009 Aug;76(2):283-288. [doi: 10.1016/j.pec.2009.01.013] [Medline: 19286343]
40. Rashid S, Soo S, Sivaji A, Naeni HS, Bahri S. Preliminary usability testing with eye tracking and FCAT analysis on occupational safety and health websites. Procedia Soc Behav Sci 2013 Nov;97:737-744. [doi: 10.1016/j.sbspro.2013.10.295]
41. Wolpin SE, Halpenny B, Whitman G, McReynolds J, Stewart M, Lober WB, et al. Development and usability testing of a web-based cancer symptom and quality-of-life support intervention. Health Informatics J 2015 Mar;21(1):10-23. [doi: 10.1177/1460458213495744] [Medline: 24406906]
42. Cowen L, Ball LJ, Delin J. An eye movement analysis of web page usability. In: Faulkner X, Finlay J, Detienne F, editors. People and Computers XVI-Memorable Yet Invisible. London, UK: Springer; 2002:317-335.
43. Ehmke C, Wilson S. Identifying Web usability problems from eye-tracking data. In: People and Computers XXI â€“ HCIâ€¦ but not as we know it: Proceedings of HCI 2007. 2007 Presented at: The 21st British HCI Group Annual Conference; September 3-7, 2007; Lancaster, UK. [doi: 10.1145/1531294.1531311]
44. Ball MJ, Weaver C, Abbott PA. Enabling technologies promise to revitalize the role of nursing in an era of patient safety. Int J Med Inform 2003 Jan;69(1):29-38. [Medline: 12485702]
45. Moody LE, Slocumb E, Berg B, Jackson D. Electronic health records documentation in nursing: nurses' perceptions, attitudes, and preferences. Comput Inform Nurs 2004 Dec;22(6):337-344. [Medline: 15602303]
46. Poissant L, Pereira J, Tamblyn R, Kawasumi Y. The impact of electronic health records on time efficiency of physicians and nurses: a systematic review. J Am Med Inform Assoc 2005 Oct;12(5):505-516 [FREE Full text] [doi: 10.1197/jamia.M1700] [Medline: 15905487]
47. Belden JL, Grayson R, Barnes J. Healthcare Information and Management Systems Society (HIMSS). 2009. Defining and testing EMR usability: principles and proposed methods of EMR usability evaluation and rating URL: http://www.himss.org/files/HIMSSorg/content/files/himss_definingandtestingemrusability.pdf [accessed 2014-11-19] [WebCite Cache ID 6UD3dOVw4]
48. Liu C, Zhu Q, Holroyd KA, Seng EK. Status and trends of mobile-health applications for iOS devices: A developer's perspective. J Syst Softw 2011 Nov;84(11):2022-2033. [doi: 10.1016/j.jss.2011.06.049]
49. Acton T, Golden W, Gudee S, Scott M. Usability and acceptance in small-screen information systems. In: Proceedings of 9th European Collaborative Electronic Commerce Technology and Research Conference. 2004 Presented at: The 9th European Collaborative Electronic Commerce Technology and Research Conference; January 1, 2004; Surrey, UK URL: http://www.pmn.co.uk/smallscreenusability.pdf
50. He J, Siu C, Chaparro B, Strohl J. Mobile. In: Bergstrom J, Schall A, editors. Eye Tracking in User Experience Design. Waltham, MA: Morgan Kaufmann; 2014:255-290.
51. Virzi RA, Sokolov J, Karis D. Usability problem identification using both low-high-fidelity prototypes. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY: ACM; 1996 Presented at: CHI '96; April 13-18, 1996; Vancouver, BC p. 236-243. [doi: 10.1145/238386.238516]
52. Carayon P, Schoofs Hundt A, Karsh BT, Gurses AP, Alvarado CJ, Smith M, et al. Work system design for patient safety: the SEIPS model. Qual Saf Health Care 2006 Dec;15 Suppl 1:i50-i58 [FREE Full text] [doi: 10.1136/qshc.2005.015842] [Medline: 17142610]
53. Berg M. Patient care information systems and health care work: a sociotechnical approach. Int J Med Inform 1999 Aug;55(2):87-101. [Medline: 10530825]
54. Carayon P. Human factors of complex sociotechnical systems. Appl Ergon 2006 Jul;37(4):525-535. [doi: 10.1016/j.apergo.2006.04.011] [Medline: 16756937]

55. Rivera-Rodriguez AJ, Karsh BT. Interruptions and distractions in healthcare: review and reappraisal. Qual Saf Health Care 2010 Aug;19(4):304-312. [doi: 10.1136/qshc.2009.033282] [Medline: 20378621]
56. Yang Y, Rivera AJ, Fortier C, Abernathy J. A macroergonomic usability evaluation of pre-filled medication syringes. In: Proceedings of the International Symposium of Human Factors and Ergonomics in Healthcare. 2014 Presented at: The International Symposium of Human Factors and Ergonomics in Healthcare; March 16-19, 2014; Chicago, IL p. 87-91. [doi: 10.1177/2327857914031013]
57. Meingast M, Roosta T, Sastry S. Security and privacy issues with health care information technology. In: Proceedings of the 28th Annual International Conference of the IEEE: Engineering in Medicine and Biology Society, 2006. EMBS '06.: IEEE; 2006 Presented at: 28th Annual International Conference of the IEEE: Engineering in Medicine and Biology Society, 2006; August 30-September 3, 2006; New York, NY p. 5453-5458 URL: http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=4463039&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D4463039 [doi: 10.1109/IEMBS.2006.260060]
58. Thomas J, Kellogg W. Minimizing ecological gaps in interface design. IEEE Softw 1989 Jan;6(1):78-86. [doi: 10.1109/52.16905]
59. Olmsted-Hawala E, Holland TH, Quach V. Usability testing. In: Bergstrom J, Schall A, editors. Eye Tracking in User Experience Design. Waltham, MA: Morgan Kaufmann; 2014:49-80.
60. Henderson J. Human gaze control during real-world scene perception. Trends Cogn Sci 2003 Nov;7(11):498-504. [Medline: 14585447]
61. Van Someren MW, Barnard YF, Sandberg JAC. The Think Aloud Method: A Practical Guide to Modelling Cognitive Processes. Volume 2. London, UK: Academic Press; 1994.
62. Hertzum M, Holmegaard KD. Thinking aloud influences perceived time. Hum Factors 2015 Feb;57(1):101-109. [Medline: 25790572]
63. Bavelas JB, Coates L, Johnson T. Listener responses as a collaborative process: the role of gaze. J Commun 2002 Sep;52(3):566-580. [doi: 10.1111/j.1460-2466.2002.tb02562.x]
64. Elling S, Lentz L, de Jong M. Retrospective think-aloud method: using eye movements as an extra cue for participants verbalizations. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY: ACM; 2011 Presented at: CHI '11; May 7-12, 2011; Vancouver, BC p. 1161-1170 URL: http://dl.acm.org/citation.cfm?id=1979116 [doi: 10.1145/1978942.1979116]
65. Guan Z, Lee S, Cuddihy E, Ramey J. The validity of the stimulated retrospective think-aloud method as measured by eye tracking. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY: ACM; 2006 Presented at: CHI '06; April 22-27, 2006; MontrÃ©al, QC p. 1253-1262 URL: http://dl.acm.org/citation.cfm?id=1124961&CFID=498086454&CFTOKEN=79430137 [doi: 10.1145/1124772.1124961]

Abbreviations

CPOE: computerized provider order entry
DCIDA: Dynamic Computer Interactive Decision Application
ED: emergency department
EHR: electronic health record
ESRAC 2.0: Electronic Self-Report Assessment for Cancer-II
FCAT: Feedback Capture After Task
HIT: health information technology
ICU: intensive care unit
ISO: International Organization for Standardization
NASA-TLX: National Aeronautics and Space Administration Task Load Index
OSH: Occupational Safety and Health
RTA: retrospective think aloud
SMI: Sensomotoric Instruments
SPASA: Short Post-Assessment Situational Awareness
SUS: System Usability Scale

Edited by G Eysenbach; submitted 20.11.14; peer-reviewed by P Henneman, D Barkana; comments to author 11.12.14; revised version received 23.12.14; accepted 01.03.15; published 14.04.15.

Please cite as:
Asan O, Yang Y
Using Eye Trackers for Usability Evaluation of Health Information Technology: A Systematic Literature Review
JMIR Human Factors 2015;2(1):e5
URL: http://humanfactors.jmir.org/2015/1/e5/
doi:10.2196/humanfactors.4062
PMID:27026079

Original Paper

A Visualization Tool to Analyse Usage of Web-Based Interventions: The Example of Positive Online Weight Reduction (POWeR)

Emily Julia Arden-Close[1], PhD; Emily Smith[2], PhD; Katherine Bradbury[2], PhD; Leanne Morrison[2], PhD; Laura Dennison[2], PhD; Danius Michaelides[3], PhD; Lucy Yardley[2], PhD

[1]Department of Psychology, Faculty of Science and Technology, Bournemouth University, Poole, United Kingdom

[2]Academic Unit of Psychology, University of Southampton, Southampton, United Kingdom

[3]Department of Electronics and Computer Science, University of Southampton, Southampton, United Kingdom

Corresponding Author:
Emily Julia Arden-Close, PhD
Department of Psychology
Faculty of Science and Technology
Bournemouth University
Talbot Campus
Fern Barrow
Poole, BH12 5BB
United Kingdom
Phone: 44 1202 965529
Fax: 44 1202 962736
Email: eardenclose@bournemouth.ac.uk

Abstract

Background: Attrition is a significant problem in Web-based interventions. Consequently, this research aims to identify the relation between Web usage and benefit from such interventions. A visualization tool has been developed that enables researchers to more easily examine large datasets on intervention usage that can be difficult to make sense of using traditional descriptive or statistical techniques alone.

Objective: This paper demonstrates how the visualization tool was used to explore patterns in participants' use of a Web-based weight management intervention, termed "positive online weight reduction (POWeR)." We also demonstrate how the visualization tool can be used to perform subsequent statistical analyses of the association between usage patterns, participant characteristics, and intervention outcome.

Methods: The visualization tool was used to analyze data from 132 participants who had accessed at least one session of the POWeR intervention.

Results: There was a drop in usage of optional sessions after participants had accessed the initial, core POWeR sessions, but many users nevertheless continued to complete goal and weight reviews. The POWeR tools relating to the food diary and steps diary were reused most often. Differences in participant characteristics and usage of other intervention components were identified between participants who did and did not choose to access optional POWeR sessions (in addition to the initial core sessions) or reuse the food and steps diaries. Reuse of the steps diary and the getting support tools was associated with greater weight loss.

Conclusions: The visualization tool provided a quick and efficient method for exploring patterns of Web usage, which enabled further analyses of whether different usage patterns were associated with participant characteristics or differences in intervention outcome. Further usage of visualization techniques is recommended to (1) make sense of large datasets more quickly and efficiently; (2) determine the likely active ingredients in Web-based interventions, and thereby enhance the benefit they may provide; and (3) guide in designing (or redesigning) of future interventions to promote greater use and engagement by enabling users to easily access valued intervention content/tools.

Trial Registration: International Standard Randomized Controlled Trial Number (ISRCTN): 31685626; http://www.isrctn.com/ISRCTN31685626 (Archived by WebCite at http://www.webcitation.org/6YXYIw9vc).

(JMIR Human Factors 2015;2(1):e8) doi:10.2196/humanfactors.4310

KEYWORDS

data visualisations; usage; Web-based interventions

Introduction

Web-based interventions for weight management (weight loss or maintenance) have grown in popularity in recent years. There is evidence that such interventions lead to meaningful weight loss [1], particularly relative to no-intervention control groups or minimal interventions [2]. However, attrition is typically high in Web-based interventions [3-5].

In any longitudinal eHealth study, there are two different types of attrition, namely, dropout attrition, or losing participants to follow-up; and nonusage attrition (not using the intervention or low usage of the intervention). Determining nonusage and dropout attrition is an essential part of analysis of Web-based interventions, as the attrition curve may indicate the underlying cause of attrition [3]. For example, there may be steady attrition, with a consistent proportion of users discontinuing usage. Alternatively, there may be an initial phase where usage is high, followed by rapid attrition, after which a stable group of regular users remains. Further, even among regular users, some Web pages are used by almost all users who log on to the website, whereas others are never used. Although higher use of website features may be associated with weight loss, it is not clear which features improve this effect or reduce attrition [5]. It is also possible that not all users may need to complete an Internet intervention to obtain positive results—different doses may be necessary for different people [6].

Several recent studies have attempted to identify the relationship between Web usage and benefit from weight management interventions. For example, Funk and colleagues [7] categorized users of a Web-based weight loss intervention as having "consistent usage," "some usage," or "minimal usage." The mean weight change was significantly higher in the "consistent usage" category, and significantly more consistent users maintained clinically important weight loss than those in the other groups. Within Internet interventions, more logins, weight and exercise entries, and use of additional features of the website after weight entry have been associated with better weight outcomes [7,8]. More specifically, use of website feedback features, such as progress charts, have been shown to be the best predictors of initial 6-month weight loss, whereas social support features, such as Web chats and participant profiles, have been related to weight maintenance at 12 months [9]. Recently, greater use of a weight tracker was associated with greater weight loss [10]. However, no study has assessed in detail whether certain Web pages are more frequently used than others, or whether certain groups of people are more likely to use particular pages. This would enable researchers to refine the content of their Web-based interventions, for example, to enable easier access to the most useful Web pages, or encourage greater use of useful but underused Web pages by identifying and addressing barriers to usage.

Positive Online Weight Reduction [11] was developed as a Web-based weight management intervention for use in primary care that aimed to result in sustainable weight loss. It was tested in a feasibility trial that consisted of 4 groups, namely: Web only, Web plus basic nurse support, Web plus regular nurse support, and usual care, to assess the extent to which weight loss was maintained at 12-month follow-up. It was designed to provide support for self-management of weight based on either a low-calorie or low-carbohydrate eating plan. Analysis revealed that average website usage, defined as duration of page viewing, was similar across the intervention arms, but extremely variable within groups. Although participants completed a mean of nine goal and weight reviews, this ranged from none to 43 completed during the 12-month trial.

Usage log data have been used to examine the relationship(s) between use of specific intervention components and subsequent outcomes/effectiveness [12-14]. Such analyses can reveal useful insights about the impact and relevance of particular components over the time course of an intervention. However, such analyses typically rely on making a priori assumptions about the specific intervention components that are expected to have an effect on uptake, adherence, or outcomes. By contrast, visualizations use aspects of exploratory sequencing techniques to summarize and plot the participant's usage of every intervention component over time [15].Using visual analysis allows differences in usage to emerge from the data and ensures that unanticipated relationships between usage and outcomes are not overlooked. Freely available visualization tools have been developed and argued to be useful for detecting patterns of usage and how they vary across individuals/groups; detecting usability or content issues, and thereby enable researchers to edit content for use in future Web-based interventions; and performing exploratory analysis to support the design of statistical queries to summarize data regarding whether use of particular pages is related to benefit [15].

Existing visualization tools provide a useful means to explore each individual participant's usage of an intervention, or particular aspects of all participants' usage of an intervention (such as days/dates of logins, start and end points of each login). However, to our knowledge, these tools do not allow for a detailed comparison of how all components of an intervention have been used by all participants within one sequence plot. Our research team has therefore developed a visualization tool to examine each individual participant's temporal usage of a Web-based intervention by illustrating what pages they have viewed, for how long, and in what order. Usage sequences for each individual are stacked within one visualization plot to facilitate comparison across all participants. This makes analysis quicker and easier compared with standard data analysis.

This paper first describes how the visualization tool works. We then illustrate the insights the visualization tool can provide by a detailed analysis of usage of the Positive Online Weight

Reduction (POWeR) intervention. This analysis had three main aims, which were realized using the visualization tool.

- Examine patterns of Web usage to identify the following:
  - At what point usage of POWeR drops off;
  - Whether participants accessed both the core and optional contents of the intervention; and
  - What information, advice, and tools were reused after their initial presentation;
- Carry out a moderator analysis of patient characteristics related to Web usage; and
- Determine whether usage of specific intervention pages and sections were related to weight change.

Methods

Design

As reported elsewhere [11], a randomized nonblinded feasibility trial of a Web-based weight management intervention (POWeR) for obese patients in primary care was used to compare 4 parallel groups: usual care, website only, website with basic nurse support, and website with regular nurse support. The trial was approved by the UK National Health Service National Research Ethics Service, and was registered with Current Controlled Trials (ISRCTN 31685626).

Participants and Procedures

Participants were recruited between May 2011 and December 2012 from five general practices in southern England. Inclusion criteria included being aged over 18, and having a body mass index (BMI)of 30 or more (or 28 with hypertension, hypercholesterolemia, or diabetes) documented in medical records. Exclusion criteria included being pregnant or breastfeeding, having current major mental or physical health problems, or self-reported inability to walk 100 m. Participants were followed up at 6 months and 1 year.

Intervention

The POWeR intervention [11] consisted of 12 weekly online sessions, in which users were taught active cognitive and behavioral self-regulation techniques (eg, POWeR Tools) and provided with evidence for their effectiveness and examples of how other users had successfully used them. The sessions did not differ between groups. Session 1 provided an overview of the intervention, advice on choosing the low-calorie or low-carbohydrate eating plan, helped users to set eating goals and plan how to implement them, asked users to identify personal reasons for losing weight, and explained how to use weekly weighing as a form of self-monitoring. All subsequent sessions began by asking the users to enter their current weight and report how often they had achieved each of the goals set the previous week (goal and weight review). Following this, users received automated advice based on their progress, and were able to set new goals and plans. This advice did not differ between groups. Session 2 covered getting support from the website (eg, setting automated motivational messages), friends and family, and the nurse. Session 3 helped users choose and implement a physical activity plan (walking or mixed physical activity). Sessions 1-3 were defined as core sessions, and became available weekly in sequence. After completing the first three sessions, users could then choose any one optional session each week after their goal and weight reviews from the following selection: cravings, slipups, stretching physical activity, tough times (emotional eating), busy lives (eating when busy), setting up your environment (environment restructuring), alcoholic and nonalcoholic drinks, eating out, and maintaining weight loss. The final session was a review. In addition to the new weekly sessions, users could also reaccess content from previous sessions at any time via the main home page, using either their POWeR Tools or a weight graph plotting their progress.

Data Collection and Analysis

All data were stored using the LifeGuide intervention authoring software [16], an online software that enables researchers to create Web-based interventions. This software automatically captures data regarding all Web pages accessed, and length of time spent viewing each Web page. A visualization tool was created using R to enable us to determine patterns of Web usage. The tool enables researchers to visually compare when particular parts of the intervention were viewed, for how long, and in what order, across all participants. A Web-based interface for the visualization tool was developed using the Shiny Web application for R (Figure 1). A user guide for the visualization tool will be made available shortly, and both the tool and user manual will be made available free of charge via the LifeGuide website.

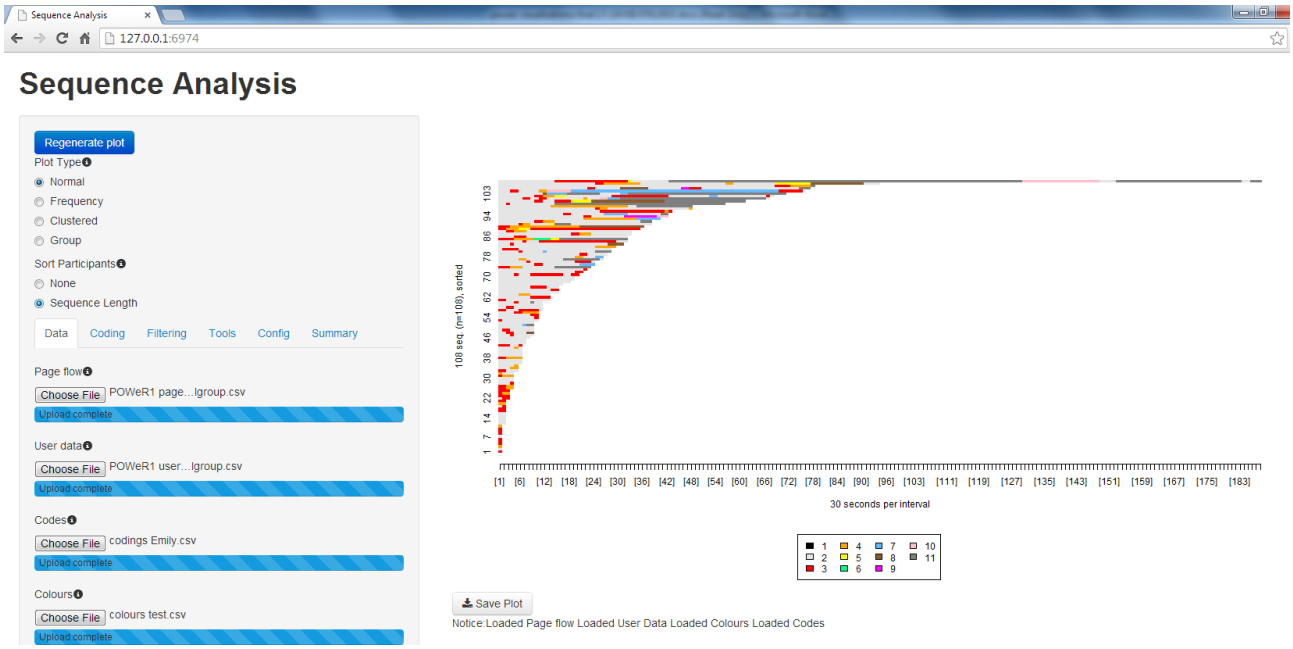
In brief, to run the tool, one needs to feed it 4 types of files: a page flow file (which shows the order in which participants have looked at pages and the time they have spent on them); a user data file (which contains data on participant characteristics and outcomes or data participants have entered into the intervention), a coding file (which assigns each intervention page a numerical code), and a color file (which assigns each intervention page code a specific color). At the top of all the interfaces, there is the option to sort participants by sequence length (the amount of time a participant has spent viewing the intervention) and choose the type of visualization plot (Table 1) the viewer wants to see.

Table 1. Different types of plots shown in a visualization.

Plot type	What it shows
Normal	Default option, shows which pages were viewed by each individual participant, in which order
Frequency	Shows usage by all participants by groups of pages, so the researcher can see which groups of pages are most used
Clustered	Groups participants into statistically similar usage patterns
Group	Allows you to see two or more visualizations next to each other, split into different types of usage patterns or users

The visualization can be filtered based on variables in the dataset (eg, user characteristics or outcomes) or which groups of pages users have/have not seen. If you have run a visualization that you want to follow-up on through statistical analysis, the tool can create an Excel file (Microsoft, Redmond, WA, USA) that lists details of all users who have seen a particular group of intervention pages.

Figure 1. Screenshot of the visualization tool.



Statistical Data Analysis

Data analysis for the moderators (use of the optional sessions, food diary, and steps diary) was carried out using SPSS (SPSS Inc, Chicago, IL, USA).

Results

Patterns of Web Usage

Overall, 195 participants consented to take part in the feasibility trial of POWeR; 16 were enrolled at a general practice, but never used the website, and therefore, were not randomized. Participants assigned to usual care (n=43) did not have access to the website after completing questionnaires, and their data were therefore not used. There were 4 participants who went online and were assigned to a group, but never used a session. To analyze Web usage, data from the 132 participants who had viewed at least one page of a session, which comprised the groups "Web only," "Web + basic nurse support," and "Web + regular nurse support" were included.

Participant characteristics for the overall sample are presented in Table 2. they are not broken down by group as this information is reported in the main power paper [11].

To analyze patterns of POWeR usage, we first carried out broad-level visualizations of how participants used the entire intervention and main components of interest (eg, core versus optional sessions), followed by more fine-grained visualizations of regularly used components (eg, eating plan tools) and subsequent statistical analyses.

Table 2. Baseline characteristics of participants.

Participant characteristic	Mean (SD)
Age (years)	51.56 (12.96)
Age left education (years)	17.82 (2.93)
Body mass index (kg/m^2)	35.49 (5.70)
Weight (kg)	100.66 (21.02)
Male, n (%)	46 (33.8)

Usage of the Core and Optional Sessions

Usage of the core and optional sessions is presented in Figure 2, with each color representing a separate group of pages. For example, the light green shows usage of the first part of the eating plan pages (which introduced the eating plans), and the dark gray shows usage of the support pages. The x-axis shows the length of time spent viewing each group of pages, broken down into blocks of 30 seconds. The y-axis can be thought of as a number of lines, each representing a specific participant. Participants are presented in order of how long they spent on the intervention, with those who spent less time nearer the bottom, and those who spent more time nearer the top.

It can be seen from Figure 2 that the core eating plan session (part 1, light green; and part 2, pink) was the most widely used, followed by the core sessions on "support" (session 2, dark gray) and "physical activity" (session 3, brown). Table 3 provides a precise breakdown of the proportion of participants accessing each POWeR session (core and optional). Two thirds of the participants accessed all the core sessions. However, each optional session (except the final review session, which was made compulsory) was accessed by less than 1 in 4 of the participants. A total of 30 participants (30/132, 23%) used all the core sessions but no optional sessions. Later sessions (eg, 7-11) were viewed by only 48 participants (48/132, 36%). This contrasted with an average of 8.62 (SD 10.46) goal and weight reviews per participant (range, 0-43).

Table 3. Numbers (and percentages) of participants who used the core and optional sessions.

Session number	Session description	Participants who viewed at least one page of the session n (%)
1	Eating plan part 1[a]	132 (100)
1	Eating plan part 2[a,b]	120 (90.9)
2	Support[a,c]	104 (78.7)
3	Physical activity[a,c]	90 (68.1)
4	Cravings	28 (21.2)
5	Slipups	32 (24.2)
6	Stretching physical activity	25 (18.9)
7	Tough times	21 (15.9)
8	Busy lives	19 (14.3)
9	Setting up your environment	13 (9.8)
10	Drinks	13 (9.8)
11	Eating out	24 (18.1)
12	Maintaining weight loss[d]	36 (27.2)

[a]Core sessions

[b]Eating plan part 1 and part 2 are both part of session 1. They are presented separately to show the points during session 1 at which participants dropped out.

[c]The sessions are presented in the order in which they were listed.

[d]This session was made compulsory.

To further explore patterns of drop out, we used the visualization tool to compare the proportion of participants dropping out at different points during the first session. This revealed that 100% of participants (132/132) used part 1 of session 1, 120/132 (91%) used part 2 of session 1, and 115/132 (87%) completed session 1 (reached the last page). Separate visualizations were also produced for each trial arm (Web only, Web + basic nurse support, and Web + regular nurse support), but revealed no meaningful and substantial differences in attrition between groups.

To further explore how the optional POWeR sessions were used, we filtered the visualization plots to only contain participants who accessed at least one of the optional sessions (Figure 3). This showed that after completion of the initial core sessions, 62/132 participants (47%) accessed both the goal and weight reviews (yellow) and the optional sessions (brown) whereas 58/132 participants (44%) continued to access the goal and weight reviews but not the optional sessions. Four (of 132) participants (3%) did not use either the goal and weight reviews or the optional sessions following completion of the core sessions.

In Figure 3, each color represents a separate group of pages. For example, the green shows usage of the eating plan pages, and the yellow shows usage of the goal and weight reviews pages. The x-axis shows the length of time spent viewing each group of pages, broken down into blocks of 30 seconds. The y-axis can be thought of as a number of lines, each representing a specific participant. Participants are presented in order of how long they spent on the intervention, with those who spent less time nearer the bottom, and those who spent more time nearer the top.

Figure 3 shows that 58/132 (44%) participants used the optional sessions. It also shows that the most frequently viewed pages were those relating to part 1 of the eating plan session and the goal and weight review, whereas the optional sessions and optional tools pages were not widely used.

Figure 2. Visualization of POWeR usage of sessions by all intervention participants.

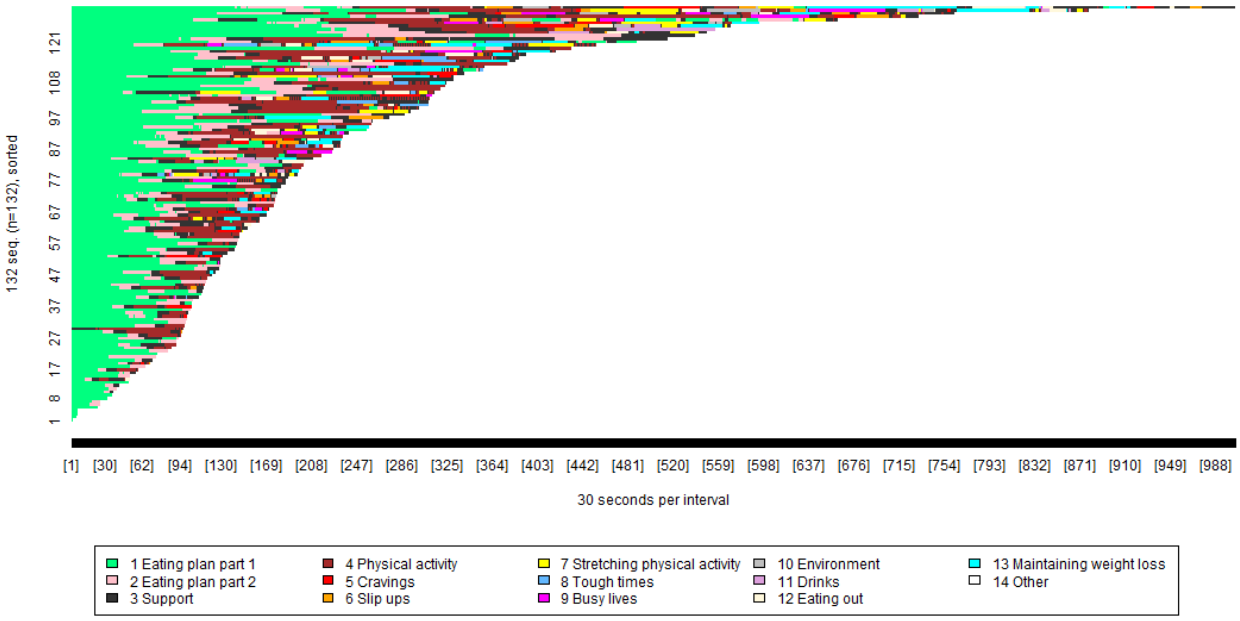
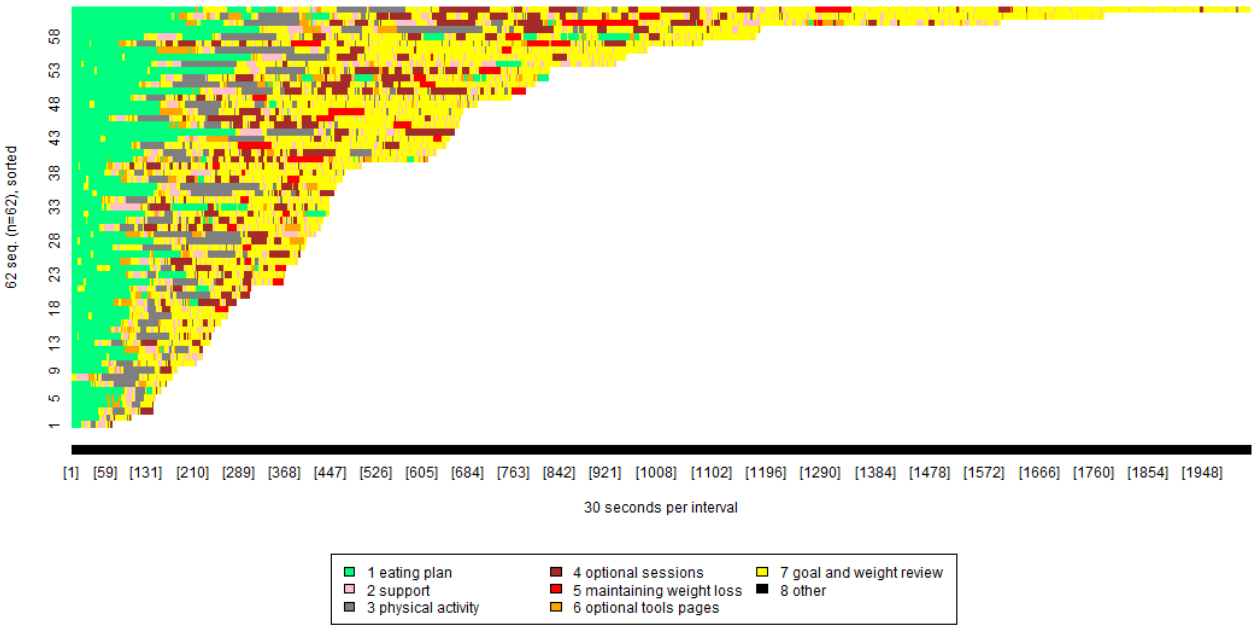


Figure 3. Visualization of POWeR usage by participants who used the optional sessions.



Repeated Use of POWeR Tools

There were 107 participants who reused at least one of the POWeR tools, as shown in Figure 4. These data are broken down as shown in Table 4.

In Figure 4, each color represents a separate group of pages. For example, the green shows usage of the eating plan pages, and the pink shows usage of the support pages. The x-axis shows the length of time spent viewing each group of pages, broken down into blocks of 30 seconds. The y-axis can be thought of as a number of lines, each representing a specific participant. Participants are presented in order of how long they spent on the intervention, with those who spent less time nearer the bottom, and those who spent more time nearer the top.

As shown in Figure 4, the POWeR tools participants reused most pages related to the eating plan (green), support (pink), and physical activity plan (dark gray). Very few participants reused the POWeR tools pages that are related to the optional sessions.

Table 4. Numbers of participants who reused POWeR tools.

Tool topic	Numbers viewed n (%)
Eating plan	91 (68.9)
Support	68 (51.5)
Physical activity plan	21 (15.9)
Slipups	7 (5.3)
Cravings	1 (0.8)
Tough times	10 (7.5)
Busy lives	7 (5.3)
Drinks	2 (1.5)
Eating out	4 (3.0)
Maintaining weight loss	17 (12.8)

We used the visualization tool to provide a detailed breakdown of the most regularly reused eating plan tools (Figure 5).

In Figure 5, each color represents a separate group of pages. For example, the pink shows usage of the weekly food diary, and the yellow shows usage of the reasons to lose weight card. The x-axis shows the length of time spent viewing each group of pages, broken down into blocks of 30 seconds. The y-axis can be thought of as a number of lines, each representing a specific participant. Participants are presented in order of how long they spent on the intervention, with those who spent less time nearer the bottom, and those who spent more time nearer the top.

As shown in Figure 5, the specific tools that appeared to be reaccessed most often were those relating to the weekly food diary (light pink), and information about eating plans (eg, lists of foods that were low/high in calories or carbohydrates—gray and dark red).

The patterns observed in Figure 5 were used to provide a more precise breakdown of the proportions of participants viewing each of the eating plan tools. This confirmed that over 40% of the participants viewed the weekly food diary (76/132, 57.5%) and information about the low-calorie (71/132, 53%) and low-carbohydrate eating plans (57/132, 43%; Table 5).

Table 5. Numbers (and percentages) of participants who reused the eating plan tools.

Eating plan topic	Code	Numbers viewed n (%)
Week 1 food diary	1	29 (21.9)
A weekly food diary	2	76 (57.5)
Low-calorie information	3	71 (53.7)
Low-carbohydrate information	4	57 (43.1)
Information about goal setting	5	9 (6.8)
Information about making plans	6	14 (10.6)
My reasons to lose weight card	7	18 (13.6)

We also used the visualization tool to provide a detailed breakdown of how the "support" tools were reused. Figure 6 shows that 68/104 participants (ie, 65% of those who were able to reaccess them) reused the tools in the "Getting Support" subcategory, which included information about the importance of getting support from others when trying to lose weight, and the various ways in which participants could get support from their nurse.

In Figure 6, each color represents a separate group of pages. For example, the light green stands for the support pages, and the pink stands for the support tools pages. The x-axis shows the length of time spent viewing each group of pages, broken down into blocks of 30 seconds. The y-axis can be thought of as a number of lines, each representing a specific participant. Participants are presented in order of how long they spent on the intervention, with those who spent less time nearer the bottom, and those who spent more time nearer the top.

This visualization shows that although some participants reused the "getting support" tools all in one go after accessing the session on "getting support," it was more common to follow each brief usage of the "getting support" session with reuse of the "getting support" tools. Table 6 provides a precise breakdown of the proportion of participants using each of the support tools.

Table 6. Numbers of participants who reused the support tools.

Support topic	Numbers viewed n (%)
Getting support	68 (65.3)
Sending motivational emails	3 (2.8)
Ask the nurse	6 (5.8)
Social times	1 (0.9)

Finally, we carried out a visualization to examine how participants reused the physical activity plan tools, as shown in Figure 7.

In Figure 7, each color represents a separate group of pages. For example, the orange shows usage of the steps diary and the light green shows usage of pages on getting more active. The x-axis shows the length of time spent viewing each group of pages, broken down into blocks of 30 seconds. The y-axis can be thought of as a number of lines, each representing a specific participant. Participants are presented in order of how long they spent on the intervention, with those who spent less time nearer the bottom, and those who spent more time nearer the top.

Figure 7 shows that the most widely reused physical activity tools pages were the steps diary (orange) and the pages on getting more active (light green), but that some of the other tools were used only by one person. Table 7 provides a precise breakdown of the proportion of participants using each of the physical activity tools.

Table 7. Numbers (and percentages) of participants who used the physical activity tools.

Physical activity topic	Numbers viewed n (%)
Getting more active	4 (3.0)
Thinking about fitting physical activity into your day	1 (0.8)
Information about the walking plan	0 (0)
Information about the mixed physical activity plan	1 (0.8)
Thinking about your walking experiences	0 (0)
Thinking about your physical activity experiences	1 (0.8)
Making a detailed walking plan	0 (0)
Making a detailed physical activity plan	1 (0.8)
Steps diary	17 (12.9)

Figure 4. Visualization of participantsâ€™ repeated use of optional tools pages.

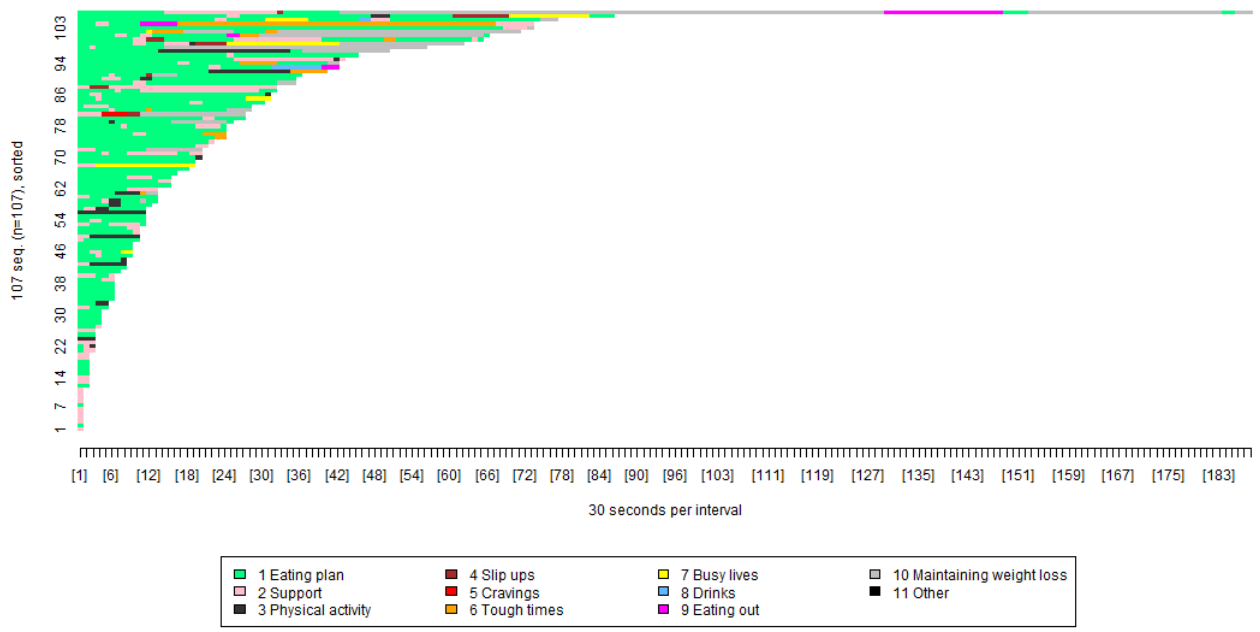


Figure 5. Visualization of participantsâ€™ repeated use of eating plan tools.

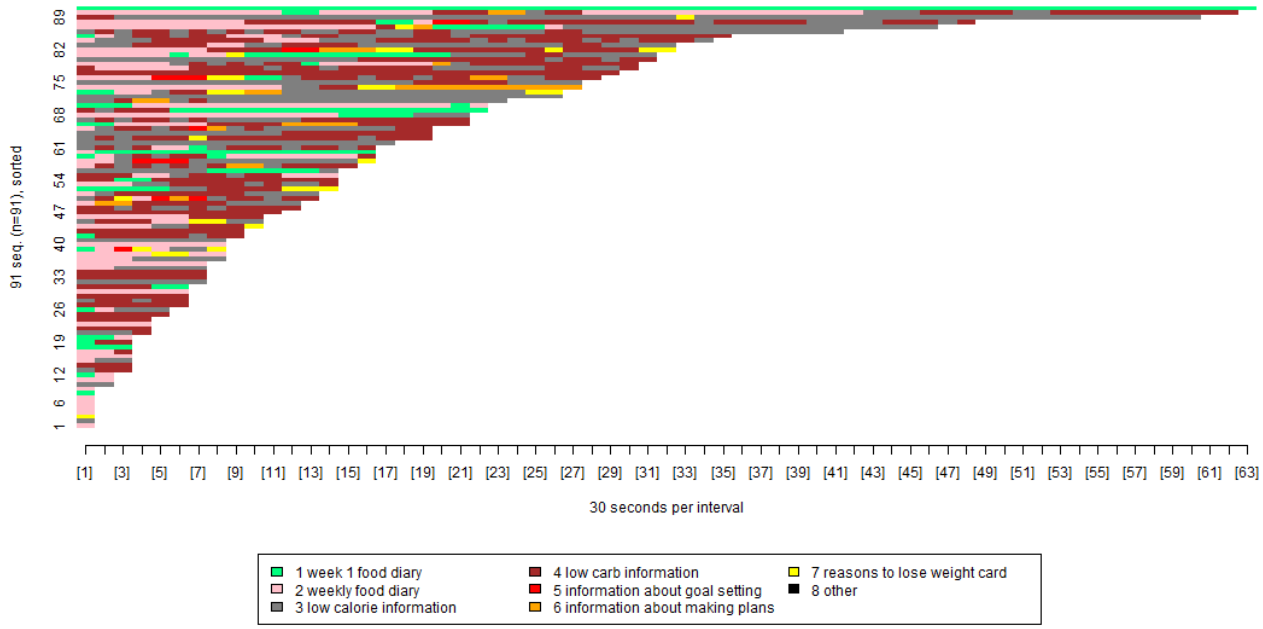


Figure 6. Visualization of reuse of the support tools in relation to the session on getting support.

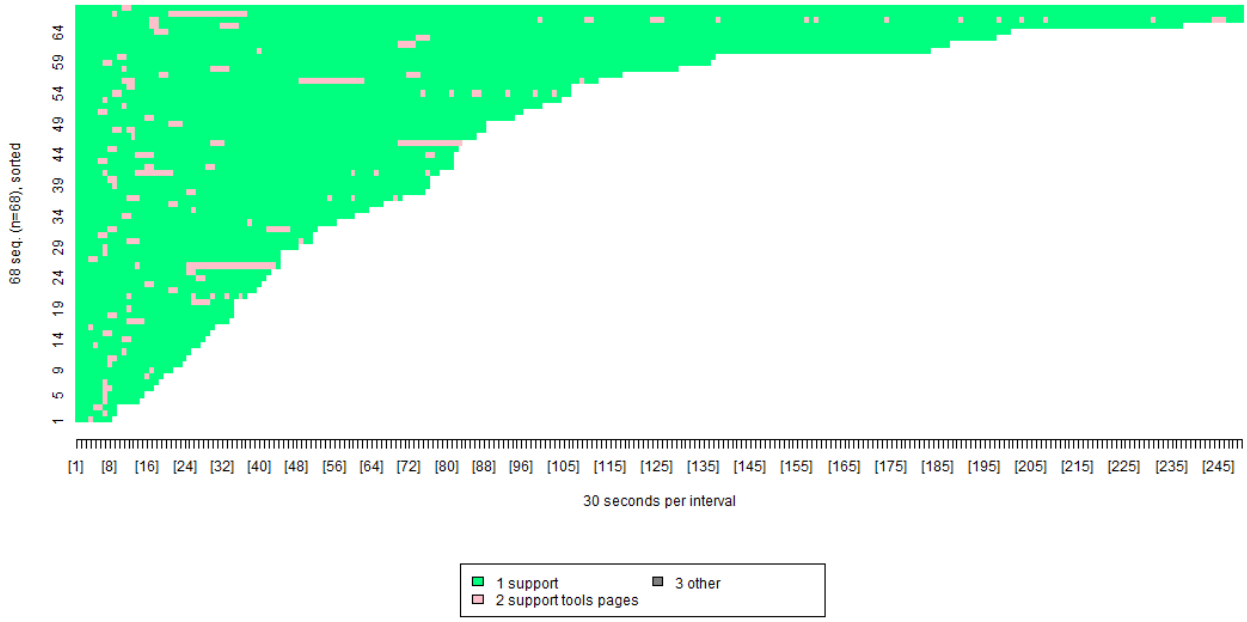
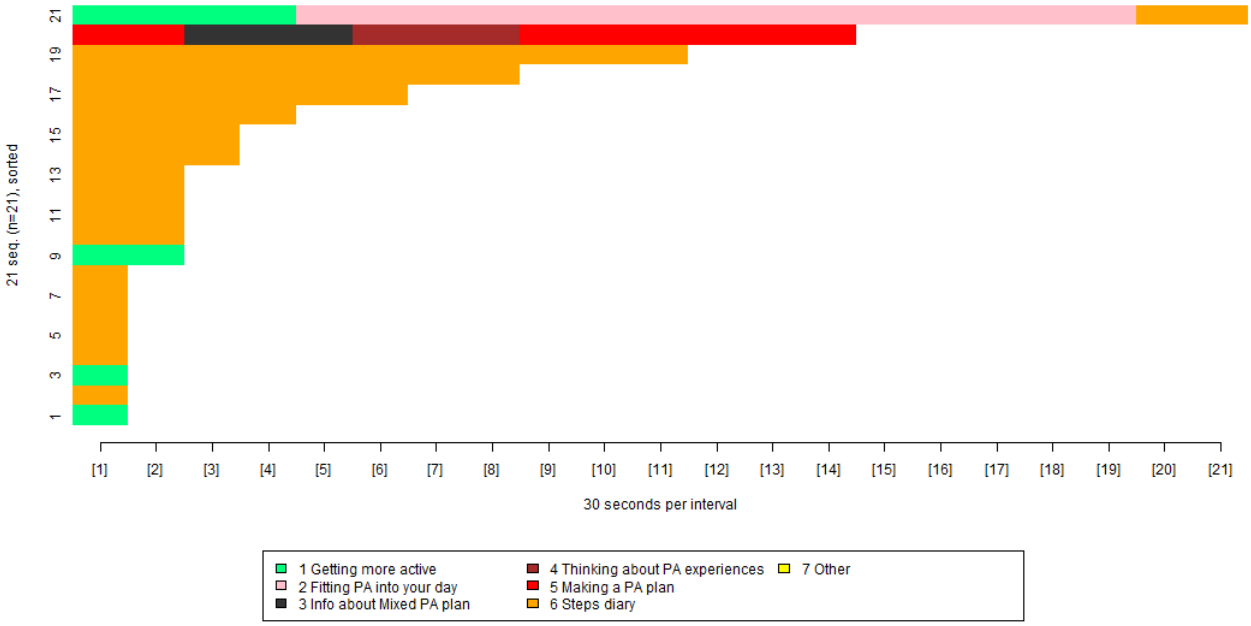


Figure 7. Visualization of participantsâ€™ repeated use of the physical activity plan tools.



Patient Characteristics Related to Web Usage

Using the visualization tool, we were able to download the IDs of participants who followed particular usage patterns. This enabled the creation of a new usage variable that detailed who had/had not used particular intervention components and could be followed up with further statistical analysis using SPSS.

Usage of Optional Sessions

Sixty-two participants used both the goal and weight reviews and the optional sessions, but 58 accessed the goal and weight reviews but not the optional sessions. Participants who did not use the optional sessions had a higher BMI at baseline (36.68 vs 34.60), were more likely to use the low-carbohydrate plan (χ^2_2=8.71, *P*=.03) and were more likely to use the walking plan (χ^2_2=2.08, *P*<.001). For these analyses, participants were classified as using the last plan they used. There was no difference in weight loss (kilograms) between those who used the optional sessions and those who did not, 3.67kg (SD 6.42) versus 2.14kg (SD 4.75; t_{134}=1.54, *P*=.13).

Repeated Use of Eating Plan Tools

Overall, 106 participants reused the eating plan tools, of whom 76 reused the weekly food diary. Participants who reused the weekly food diary were older, 53.62 versus 48.95 (t_{134}=−2.11, *P*=.04), and completed more goal and weight review sessions than those who did not reuse the diary, 8.89 versus 3.23 ($t_{125.34}$=−3.64, *P*<.001). There was no difference in weight loss between those who did and did not reuse the food diary (2.95, SD 5.53) versus (3.11, SD 6.17; t_{134}=0.16, *P*=.87).

Repeated Use of Physical Activity Tools

Overall, 21 participants reused the physical activity tools. Those who reused the steps diary were older than those who did not, 58.82 (SD 14.44) versus 50.52 (SD 12.45; t_{134}=−2.52, *P*=.01). Participants who reused the steps diary (physical activity plan tools; n=17), lost more weight than those who did not, 5.78 kg (SD 6.87) versus 2.63 kg (SD 5.56; t_{134}=−2.12, *P*=.04).

Repeated Use of Getting Support Tools

Use of getting support tools was analyzed for the nurse groups only (as the Web group did not receive nurse support). A total of 68/104 participants (65% of those who were able to reaccess them) reused the getting support tools. There were no differences at baseline between those who did and did not use the getting support tools. However, those who used the getting support tools completed more of the sessions than those who did not, 3.39 (SD 1.14) versus 0.5 (SD 0.59; $t_{77.48}$=−15.68, *P*<.001), and more goal and weight reviews than those who did not, 0 (SD 0) versus 8.81 (SD 10.65; t_{66}=−6.77, *P*<.001). They also lost more weight than those who did not, 4.03 kg (SD 6.93) versus 1.53 kg (SD 4.04; $t_{70.04}$=−2.12, *P*=.038).

Discussion

Principal Findings

This paper had three main aims, which the visualization tool was able to help us realize. These were as follows: (1) to see patterns of Web usage, (2) to carry out a moderator analysis of patient characteristics related to Web usage, and (3) to determine which pages were related to benefit from the Web-based intervention. These results are discussed in the following section in relation to these aims.

First, the visualization tool was extremely helpful in enabling us to determine patterns of Web usage. A first key observation is that the vast majority of participants who went online accessed the first session, but there was a drop of approximately 20% of participants from the first session (n=132 in part 1 and n=120 in part 2) to the second session (n=104). This is similar to the rapid attrition rate reported in similar Web-based weight loss interventions [3-5]. Dropout then continued at a rate of approximately 10% per session. Breaking down the first session

into two parts based on content covered (as it was very long and each part had a similar length to the other full sessions) and checking how many participants accessed the last page of session 1 enabled us to see that almost 90% of participants completed the first session (n=120/132). To ensure all essential information is covered, each session should be presented as early as possible in the intervention. Interventions that aim to also prioritize physical activity should present this as early on as possible.

A second key observation is that only half the participants accessed any of the optional sessions, and each optional session was viewed by less than 25% of participants. Nevertheless, nearly half the participants continued to use the weekly goal and weight review, despite deciding not to access new optional content. In retrospect, this pattern of usage could have been unintentionally prompted by the design of the page following goal review, as the logout option was prominently placed. Alternatively, it could mean that participants felt the additional sessions were neither necessary nor particularly novel (as they covered topics that are commonly addressed by other weight management interventions). In support of this interpretation, there were no differences in weight loss between participants who did and did not use the optional sessions, indicating that the optional content was indeed not necessary for weight loss. In addition, those who chose not to access the optional content had a higher BMI at baseline, so may have been more likely to have encountered similar content in previous weight management attempts. This finding justifies the decision to make these sessions optional, and also suggests that for many participants the goal and weight review (which provided individualized progress-relevant feedback messages as well as a weight loss graph) was more important and rewarding to access than the generic weight management advice.

The eating plan tools were the most reused, especially the weekly food diary, and information about the low-calorie eating plan and the low-carbohydrate eating plan. Thus, explorations of usage patterns using visualization tools can help to identify the particular intervention tools that participants are keen to reuse online. Such insights can help design hybrid interventions that enable access to selected intervention content through multiple digital devices (eg, mobile phone apps). For example, a mixed-methods evaluation of a supplemental POWeR mobile phone app also showed that participants particularly valued being able to reaccess food lists associated with their eating plans on the go via their mobile phone [17]. The eating plan tools were the most basic weight management tools, and less essential tools such as the motivational “reasons to lose weight” or “sending motivational emails” support tools were not reused. However, this does not necessarily mean that the less essential tools were not valued by participants. It could be that participants engaged with these tools at their first presentation during the core session (eg, by printing out their reasons to lose weight card or setting up support emails there and then) and did not need to reuse them via the POWeR website.

Those who reused the food diary were older and had completed more goal and weight reviews than those who had not. It is possible that these participants may have been more conscientious in their attitude to weight loss, or that younger participants could have been using alternative tools. However, it is important to note that those who reused the food diary did not lose more weight than other participants. To minimize the intrusiveness and burden of weight management, POWeR specifically encourages users to employ food diaries only occasionally, as diagnostic tools when necessary, and not to rely on them for long-term weight management[18].

Those who reused the getting support tools had completed more sessions and goal and weight reviews and lost more weight than those who did not. This suggests that the support tools were helpful in enabling weight loss. The challenge now is engaging with those users who did not use the support tools. Interestingly, very few people reused the “ask the nurse” function, which allowed users to send queries or messages to the nurse providing them with support. Some POWeR users have indicated in our follow-up interviews that they would like to be able to access human support when they feel the need [19], but it appears that the facility to send the nurse an email may not meet this need. This could be because email is an insufficiently personal medium to access support [20], but it could also indicate that the opportunity to contact the nurse should be presented differently in future interventions; for example, perhaps offered as an immediate option in goal feedback if participants are not meeting their goals (rather than requiring users to access the option from their tools). Alternatively, these findings may indicate that people did not feel the need to contact the nurse, although they felt that it was helpful to have the option there.

Very few people reused the physical activity tools, suggesting that physical activity may not have been seen as an important part of weight management by POWeR users. However, of the physical activity tools, the steps diary was the most widely reused, and was associated with greater weight loss. Users of the steps diary may have used pedometers. They may also have had increased levels of autonomous motivation as this has mediated the effect of self-monitoring and diary usage on weight loss in previous studies [21]. It may therefore be beneficial to find ways to increase repeated and regular usage of the steps diary [22]. It is important to note that participants could only reuse the steps diary if they had chosen to follow the walking plan. From these results it is therefore not clear whether it was specifically the steps diary that was useful, or whether the walking plan was more beneficial than the mixed physical activity plan.

Limitations

This study had several limitations. First, the results described here are based on a single feasibility study, and it is unclear how widely they would apply to a wider population. In particular, the sample participating in POWeR included fewer men and very few members of ethnic minorities. However, the sample was not young or highly educated, and as such could be considered broadly representative of the population eligible to enroll in such an intervention in primary care [11]. Second, although our exploratory analyses identified a number of possible patterns in Web usage and associations with outcome, further research is needed to confirm these patterns and test the hypotheses arising from this study. Third, the results regarding use of the steps diary and weight loss were based on a small

number of users of the steps diary, and should therefore be interpreted with caution. This needs to be replicated with larger populations. Fourth, we considered the intervention groups from POWeR as a single population. It is possible that nurse support may have influenced Web usage. We were not able to determine this due to the small sample size.

Conclusions

The visualization tool provided a useful and efficient method for interpreting and exploring a very large dataset on usage of a Web-based weight management intervention. Specifically, the visualization tool helped to determine aspects of the intervention design and content that seem to encourage and discourage repeated use. Insights gained from a visual analysis of usage data also helped to determine the associations between usage patterns, participant characteristics, and weight change in subsequent statistical analyses. The visualization tool complements the work of Morrison and Doherty [15] by enabling an in-depth analysis of all participants' usage of EVERY intervention component within one sequence plot. Different visualization tools are likely to be more or less useful depending on the intervention architecture and research questions of interest. The visualization tool presented here may be particularly useful for inductive analyses of tunneled interventions. By contrast, the toolkits developed by van Gemert-Pijnen and colleagues [12] may be particularly beneficial for usage analyses following a priori assumptions about key intervention components. The toolkit developed by Morrison and Doherty[15] may be particularly beneficial for individual-level analyses or group-level analyses of nontunneled interventions that do not have a clear start and end point. Visualization toolkits can be used as part of a mixed-methods approach for developing and evaluating digital interventions that seek to arrive at a more complete picture of the differences in the way in which participants use an intervention, supplemented by qualitative insights about participants' subjective experiences of using the intervention [23] and quantitative data on the effect of the intervention on health-related outcomes. Further usage of visualization techniques is highly recommended in order to (1) guide the design (redesign) of future interventions so that they enable easy access to valued intervention content, and (2) unlock the active ingredients in Web-based interventions, so they can be enhanced to reach and engage the maximum eligible population.

Acknowledgments

The development of the visualization tool was funded by the UK Economic and Physical Sciences Research Council (EP/I032673/1: UBhave: ubiquitous and social computing for positive behaviour change). The authors are grateful to Peter Smith, Mark Weal, Charlie Hargood, and Matthew Taylor for their contributions to its development.

Conflicts of Interest

None declared.

References

1. Neve M, Morgan PJ, Jones PR, Collins CE. Effectiveness of web-based interventions in achieving weight loss and weight loss maintenance in overweight and obese adults: A systematic review with meta-analysis. Obes Rev 2010 Apr 11;11(4):306-321. [doi: 10.1111/j.1467-789X.2009.00646.x] [Medline: 19754633]
2. Wieland LS, Falzon L, Sciamanna CN, Trudeau KJ, Brodney S, Schwartz JE, et al. Interactive computer-based interventions for weight loss or weight maintenance in overweight or obese people. Cochrane Database Syst Rev 2012;8:CD007675 [FREE Full text] [doi: 10.1002/14651858.CD007675.pub2] [Medline: 22895964]
3. 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]
4. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JEWC. Persuasive system design does matter: A systematic review of adherence to web-based interventions. J Med Internet Res 2012;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
5. Neve MJ, Collins CE, Morgan PJ. Dropout, nonusage attrition, and pretreatment predictors of nonusage attrition in a commercial web-based weight loss program. J Med Internet Res 2010;12(4):e69 [FREE Full text] [doi: 10.2196/jmir.1640] [Medline: 21156470]
6. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. Ann Behav Med 2009 Aug;38(1):18-27 [FREE Full text] [doi: 10.1007/s12160-009-9133-4] [Medline: 19802647]
7. Funk KL, Stevens VJ, Appel LJ, Bauck A, Brantley PJ, Champagne CM, et al. Associations of internet website use with weight change in a long-term weight loss maintenance program. J Med Internet Res 2010;12(3):e29 [FREE Full text] [doi: 10.2196/jmir.1504] [Medline: 20663751]
8. Cussler EC, Teixeira PJ, Going SB, Houtkooper LB, Metcalfe LL, Blew RM, et al. Maintenance of weight loss in overweight middle-aged women through the Internet. Obesity (Silver Spring) 2008 May;16(5):1052-1060. [doi: 10.1038/oby.2008.19] [Medline: 18309301]

9. Krukowski RA, Harvey-Berino J, Ashikaga T, Thomas CS, Micco N. Internet-based weight control: The relationship between web features and weight loss. Telemed J E Health 2008 Oct;14(8):775-782 [FREE Full text] [doi: 10.1089/tmj.2007.0132] [Medline: 18954247]
10. Brindal E, Freyne J, Saunders I, Berkovsky S, Smith G, Noakes M. Features predicting weight loss in overweight or obese participants in a web-based intervention: Randomized trial. J Med Internet Res 2012;14(6):e173 [FREE Full text] [doi: 10.2196/jmir.2156] [Medline: 23234759]
11. Yardley L, Ware LJ, Smith ER, Williams S, Bradbury KJ, Arden-Close EJ, et al. Randomised controlled feasibility trial of a web-based weight management intervention with nurse support for obese patients in primary care. Int J Behav Nutr Phys Act 2014;11:67 [FREE Full text] [doi: 10.1186/1479-5868-11-67] [Medline: 24886516]
12. Van Gemert-Pijnen JE, Kelders SM, Bohlmeijer ET. Understanding the usage of content in a mental health intervention for depression: An analysis of log data. J Med Internet Res 2014;16(1):e27 [FREE Full text] [doi: 10.2196/jmir.2991] [Medline: 24486914]
13. Kim HK, Niederdeppe J, Guillory J, Graham M, Olson C, Gay G. Determinants of pregnant women's online self-regulatory activities for appropriate gestational weight gain. Health Commun 2014 Sep 10:1-11. [doi: 10.1080/10410236.2014.905900] [Medline: 25205417]
14. Hendricks MC, Ward CM, Grodin LK, Slifer KJ. Multicomponent cognitive-behavioural intervention to improve sleep in adolescents: A multiple baseline design. Behav Cogn Psychother 2014 May;42(3):368-373. [doi: 10.1017/S1352465813000623] [Medline: 25267109]
15. Morrison C, Doherty G. Analyzing engagement in a web-based intervention platform through visualizing log-data. J Med Internet Res 2014;16(11):e252 [FREE Full text] [doi: 10.2196/jmir.3575] [Medline: 25406097]
16. LifeGuide. 2015. URL: https://www.lifeguideonline.org/ [accessed 2015-02-02] [WebCite Cache ID 6W2u97RYz]
17. Morrison LG, Hargood C, Lin SX, Dennison L, Joseph J, Hughes S, et al. Understanding usage of a hybrid website and smartphone app for weight management: A mixed-methods study. J Med Internet Res 2014;16(10):e201 [FREE Full text] [doi: 10.2196/jmir.3579] [Medline: 25355131]
18. Yardley L, Williams S, Bradbury K, Garip G, Renouf S, Ware L, et al. Integrating user perspectives into the development of a web-based weight management intervention. Clin Obes 2012 Oct;2(5-6):132-141. [doi: 10.1111/cob.12001] [Medline: 25586248]
19. Bradbury K, Dennison L, Little P, Yardley L. Using mixed methods to develop and evaluate an online weight management intervention. Br J Health Psychol 2015 Feb;20(1):45-55. [doi: 10.1111/bjhp.12125] [Medline: 25406436]
20. Renouf S, Bradbury K, Yardley L, Little P. The role of nurse support within an internet-delivered weight management intervention: A qualitative study. Psychol Health Med 2014 Dec 2:1-9. [doi: 10.1080/13548506.2014.986138] [Medline: 25438990]
21. Webber KH, Gabriele JM, Tate DF, Dignan MB. The effect of a motivational intervention on weight loss is moderated by level of baseline controlled motivation. Int J Behav Nutr Phys Act 2010;7:4 [FREE Full text] [doi: 10.1186/1479-5868-7-4] [Medline: 20157441]
22. Danaher BG, Seeley JR. Methodological issues in research on web-based behavioral interventions. Ann Behav Med 2009 Aug;38(1):28-39 [FREE Full text] [doi: 10.1007/s12160-009-9129-0] [Medline: 19806416]
23. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: Application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]

Abbreviations

BMI: body mass index
POWeR: Positive Online Weight Reduction

Edited by G Eysenbach; submitted 03.02.15; peer-reviewed by C Morrison, K Funk; comments to author 23.02.15; revised version received 24.03.15; accepted 14.04.15; published 19.05.15.

Please cite as:
Arden-Close EJ, Smith E, Bradbury K, Morrison L, Dennison L, Michaelides D, Yardley L
A Visualization Tool to Analyse Usage of Web-Based Interventions: The Example of Positive Online Weight Reduction (POWeR)
JMIR Human Factors 2015;2(1):e8
URL: http://humanfactors.jmir.org/2015/1/e8/
doi:10.2196/humanfactors.4310
PMID:27026372

Original Paper

Evaluation of Home Health Care Devices: Remote Usability Assessment

Philip Kortum[1*], PhD; S Camille Peres[2*], PhD

[1]Rice University, Department of Psychology, Houston, TX, United States

[2]Texas A&M University Health Science Center, Department of Environmental and Occupational Health, College Station, TX, United States

[*]all authors contributed equally

Corresponding Author:
Philip Kortum, PhD
Rice University
Department of Psychology
6100 Main Street MS25
Houston, TX, 77005
United States
Phone: 1 7133484813
Fax: 1 7133485221
Email: pkortum@rice.edu

Abstract

Background: An increasing amount of health care is now performed in a home setting, away from the hospital. While there is growing anecdotal evidence about the difficulty patients and caregivers have using increasingly complex health care devices in the home, there has been little systematic scientific study to quantify the global nature of home health care device usability in the field. Research has tended to focus on a handful of devices, making it difficult to gain a broad view of the usability of home-care devices in general.

Objective: The objective of this paper is to describe a remote usability assessment method using the System Usability Scale (SUS), and to report on the usability of a broad range of health care devices using this metric.

Methods: A total of 271 participants selected and rated up to 10 home health care devices of their choice using the SUS, which scores usability from 0 (unusable) to 100 (highly usable). Participants rated a total of 455 devices in their own home without an experimenter present.

Results: Usability scores ranged from 98 (oxygen masks) to 59 (home hormone test kits). An analysis conducted on devices that had at least 10 ratings showed that the effect of device on SUS scores was significant ($P<.001$), and that the usability of these devices was on the low end when compared with other commonly used items in the home, such as microwave ovens and telephones.

Conclusions: A large database of usability scores for home health care devices collected using this remote methodology would be beneficial for physicians, patients, and their caregivers.

(JMIR Human Factors 2015;2(1):e10) doi:10.2196/humanfactors.4570

KEYWORDS

health care evaluation mechanisms; human-computer interaction design and evaluation methods; patient satisfaction; usability testing

Introduction

Overview

The usability of technology can be important in the consumer domain because it can drive adoption and create consumer loyalty [1]. In the medical domain, however, lack of usability can cost lives. In 2000, the Institute of Medicine (IOM) published its seminal report, *To Err is Human: Building a Safer Health System*, indicating that over 98,000 lives were lost every year in the United States alone due to preventable human errors [2]. Recent reports suggest that this number may have grown to over 400,000 [3]. Although the IOM report focused on the errors that were occurring in hospitals, they noted that "…as more care shifts to ambulatory and home settings, the use of medical technology by non-health professionals can be expected to take on increasing importance" [2] (p 63). Indeed, if some

of the most highly trained medical professionals in the world are making errors in the treatment of patients, there should be great concern in the ability of patients and their care providers to render medical treatment at home without error.

Challenges for home care are rising for multiple reasons. First, an increasing amount of health care is now done in a home setting, away from the hospital. From 1995 to 2008, there was a fivefold increase in the number of patients who received home health care from Medicaid, with an estimated 12 million people receiving some form of home health care [4]. Second, this increase in home health care is being accompanied by ever increasing levels of technology being used in the home. Third, the individuals who are expected to use this technology are likely to be minimally trained, working under stressful conditions, and may be suffering from age-related declines in cognitive, perceptual, and physical abilities—circumstances that can lead to the potential for errors, often with significant consequences [5].

The United States Food and Drug Administration (FDA) has also recognized that home health care devices are of growing importance and concern with the launching of their Medical Device Home Use Initiative in 2010 [6]. As part of that initiative, the organization has acknowledged that there are many benefits of using health care devices in the home, including cost savings resulting from fewer hospitalization days and the potential for improvements in the quality of life certain patients may enjoy because they are in a familiar and convenient venue as they receive their care. However, they also note that there are numerous usability issues surrounding the use of such devices and that these issues need to be systematically addressed.

Often times the difficulty in home-care device use stems from the fact that devices that have been designed and certified for professional medical users are then directly transferred to the home environment with little regard to the difficulty this might pose. These kinds of transitions from hospital use to patient/caregiver use might be more successful if it were acknowledged that these 2 user populations are different, and have different needs, and then these differences could be accounted for in the design process or during the development of training material. An excellent case in point is the migration of defibrillator technology from the sole domain of trained medical professional use to use by a completely untrained general population. In a study of hospital-grade defibrillators, experienced emergency medical services personnel made errors (such as trying to defibrillate before the device was ready and performing a cardioversion when they intended to defibrillate), which could cause harm to the patient [7]. This would seem to suggest that migration of this kind of mission-critical device to public use would be ill advised. However, after significant user-centered design work on the development of automatic external defibrillators (AEDs), studies have shown that untrained 6th grade school children's performance with the device was comparable with that of professional paramedics [8]. This success suggests that with proper care, even complex medical devices can be made safe and effective for use by relatively untrained individuals.

Background

For some time now, the anecdotal evidence about the difficulty of ease of use for home health care devices has been building. However, there has been little systematic scientific study to quantify the global nature of the home health care device usability problem and characterize device usability in a field situation. Much of the available literature has tended to focus around a handful of devices, such as pregnancy test kits [9], cholesterol test kits [10], glucometers, and other diabetes management tools [11]. Studies typically assess a few medical devices of a single type in a laboratory setting, making it difficult to compare usability across studies and devices. It also makes the pace of adding new usability information about specific devices exceptionally slow.

There is a growing consensus that the usability of home-care devices warrants significant additional attention. The Association for the Advancement of Medical Instrumentation (AAMI) recently released standards [12] that address the human factors requirements for highly usable medical devices and the US FDA has begun to enforce the application of these standards in the approval process of new devices. Numerous groups, including AAMI, the US FDA, the National Academies of Science, and the Human Factors and Ergonomics Society have held numerous forums, panels, and workshops aimed directly at the human factors issues associated with home health care, with the goal of highlighting the importance of the problem and to disseminate the latest research findings.

One of the key pieces of information that is currently lacking in this domain is a quantitative assessment of the usability of a broad range of home health care devices. Designers, physicians, home health nurses, caregivers, and patients would all benefit by having a better understanding of how usable (or unusable) different home health care technologies really are. Physicians could use the information to make more informed decisions about what kinds of home health care might be appropriate for their patient, particularly those who might have physical or cognitive declines. Home health care nurses could use the information to determine what devices might need extra attention when showing a household member how to use that device. Patients and family caregivers could use the information to help select home health care devices that had the best usability profiles. Further, patient compliance and adherence to medical advice is a known issue [13] and patients and caregivers are much more likely to adopt and use medical devices if they believe that those devices will be easy to use [14].

Indeed, poor home health care device usability made it to the ECRI Institute's top 10 health technology hazards of 2012 [15]. One of their recommendations was for doctors to consider the usability of the devices they were going to prescribe for their patients. However, this information does not currently exist for the wide variety of home health care devices currently being used.

There are a number of ways that this kind of usability data could be collected. Traditional user testing is one important way. Traditional user testing takes place in a laboratory and involves bringing in representative users, giving them a task to perform, and observing their performance as they try to accomplish the

task on the given product or service. The International Standards Organization usability metrics [16] of effectiveness (accuracy and completion of tasks), efficiency (time on task, physical or mental effort, rate of throughput), and satisfaction are generally collected and used to assess the usability of the product. This kind of testing could also take place in a home or hospital setting or it could take place remotely, with the experimenter conducting the test from a distant location, while the patient uses a device in the home. Other evaluation methods, often described as discount-usability techniques, could also be employed. In these methods, experts make assessments of the product or service (without benefit of real users) by employing a set of usability heuristics and determining how well the product conforms to those heuristics. The difficulty with these traditional methods is that they require extensive time to perform, and so the number of devices that can be evaluated, as well as the number of users who can evaluate each device, is greatly limited.

In this paper, we describe a very different method of collecting data—a remote usability assessment method using a survey that captures a user's assessment of the usability of a product or service, or in this case, a home health care device. The advantage of using such a method is that it allows for a much broader and larger sample, and eliminates some of the issues associated with small usability samples [17]. More users from diverse groups can be assessed and a greater number of devices can be evaluated than with traditional methods. More importantly, users can base their usability assessments on the totality of all their experiences with the device, rather than a single in-laboratory interaction. The method can be applied to a specific brand and model of device (eg, Acme Glucometer Model X-123) or to a class of devices, without regard to a specific model or brand (eg, glucometers). Collecting data on classes of products allows researchers to make more generalizable assessments of products that might have usability difficulties due to the nature of the task they perform, or the technology required to perform that task. While there are undoubtedly differences in the usability of specific products within a class, it has been shown that the variance of the usability scores for classes of items is the same as the variance observed for usability scores of a specific item [18]. This suggests that there is general agreement about the average usability of a class of items. For example, it seems likely that most readers would agree that a standard touch-tone landline telephone is easier to use than a handheld global positioning system navigation system. Indeed, Kortum and Bangor [18] used this remote method to collect data on 14 different popular consumer goods (for both specific items and classes of items) for over 1000 users and found that the method produced reliable data. Further, Kortum and Peres [19] found that this method is comparable with usability testing for ordinally comparing the usability of devices or systems.

Methods

Data Collection

Usability data on home health care devices were collected in the field remotely, without direct usability testing. Using the System Usability Scale (SUS), participants were asked to rate the subjective usability of common home health care devices with which they had direct experience.

Participants

The participants were 271 undergraduate students at Rice University (Houston, TX, USA). There were 161 female participants, 109 male participants, and 1 who responded as "other" to gender, with an average age of 19.5. Participants self-selected into the study, and were not recruited or screened on the basis of having any specific health issues.

Measures

In this study, we used the SUS to assess subjective usability. The SUS is a 10-item survey instrument developed by Brooke [20] as an efficient method of determining the usability of a given product or service. There are a large number of other surveys available that also measure usability (see [21] for a review), but the SUS was chosen because it has 5 attributes that make it ideal for use in this study. First, the survey has demonstrated that it can be used to assess nearly any technology, so any number of different devices or interfaces can be assessed with the same instrument [21]. While many of these evaluated technologies have been consumer-grade systems, the SUS has also been used successfully in the medical domain for devices as diverse as insulin pumps, heart rate monitors, and glucose-monitoring devices. Second, the SUS has high reliability and has been used in a large number of studies, and therefore, its properties are well-known, with well-established benchmarks for comparative analysis [21-23]. Third, because of its short length, it can be quickly and easily administered. Fourth, the survey returns a scored value between 0 (unusable) and 100 (highly usable), which makes the interpretation of the scores easier for experts and nonexperts alike. Research relating these scores to easily understandable adjective ratings has made the interpretation of the scores even easier [24]. Finally, because the instrument is nonproprietary, it is a cost-effective choice for researchers to use.

In this study, we used the modified version of the SUS described by Bangor and colleagues [21]. This version differs from the original version of the SUS with a simple modification of question 8 (changing the word "cumbersome" to "awkward") to increase its understandability for a broader range of raters. The SUS was further modified by changing the word "system" to "medical device" to assist the user in making accurate ratings. This type of change has been demonstrated to have no impact on the validity or reliability of the survey instrument [23].

Procedure

Upon signing up for the study, participants were directed to a website that contained the survey. After completing an Institutional Review Board-approved consent form, they were queried about basic demographic information and given general instructions that described the rating task and provided exemplars of the kinds of home health care devices that were of interest. They then selected a home health care device that they had used from a list (Table 1), which was a subset of home health care devices described by Story [25]. Because our sample population comprised relatively healthy students, we only used

23 of the devices listed by Story [25], excluding those devices associated with more acute care (eg, nasogastric feeding tubes, hospital beds). Following this selection, they rated that device's usability using the SUS. There was also an option for the participant to enter any other home health care device they used, and rate its usability as well. The participant continued rating until they reached 10 devices or indicated they had no more devices to rate.

Results

Devices Rated

The participants rated a total of 455 devices. Table 1 shows the specific devices rated by the participants and the frequency of the ratings, as well as the mean and standard error of the SUS scores. As seen in Table 1, the thermometer was rated by the most participants and had one of the higher average SUS scores (80.53). The highest SUS score was for the oxygen mask (95.00), but this was only rated by 2 people. There were 8 different devices that participants listed under "other" and those were stethoscope, nebulizer, allergy nasal spray, humidifier, electronic muscle stimulator, electroencephalography (EEG), intrauterine birth control device, and BAND-AID.

Table 1. Frequency of responses for each device, as well as the mean and standard error (standard error of the mean) of the System Usability Scale (SUS) for each device.

Device	N	Mean SUS[a]	Standard error of the mean
Thermometer	227	80.53	0.88
Blood pressure cuff	71	73.56	1.72
Inhaler	62	75.97	2.15
Pregnancy test kit	23	66.74	4.27
Syringe	16	67.66	4.58
Blood glucose meter	12	69.58	5.34
Epinephrine injector (EpiPen)	11	65.00	4.89
Allergy test kit	7	71.43	4.81
Drug test kit	4	68.75	11.97
Feeding tubes	2	71.25	6.25
Hormone test kit	2	58.75	16.25
Nebulizer	2	67.50	12.50
Oxygen masks	2	95.00	5.00
BAND-AID	1	—	—
Birth control: intrauterine device	1	—	—
Catheters	1	—	—
Cholesterol test kit	1	—	—
Electroencephalography (EEG)	1	—	—
Electrocardiogram monitors	1	—	—
Electronic muscle stimulator	1	—	—
Humidifier	1	—	—
Intravenous equipment	1	—	—
Nasal spray (allergy)	1	—	—
Stethoscope	1	—	—
Transcutaneous electrical nerve stimulation equipment	1	—	—
Ventilators	1	—	—

[a]The SUS scores are not reported for those devices that had only 1 response.

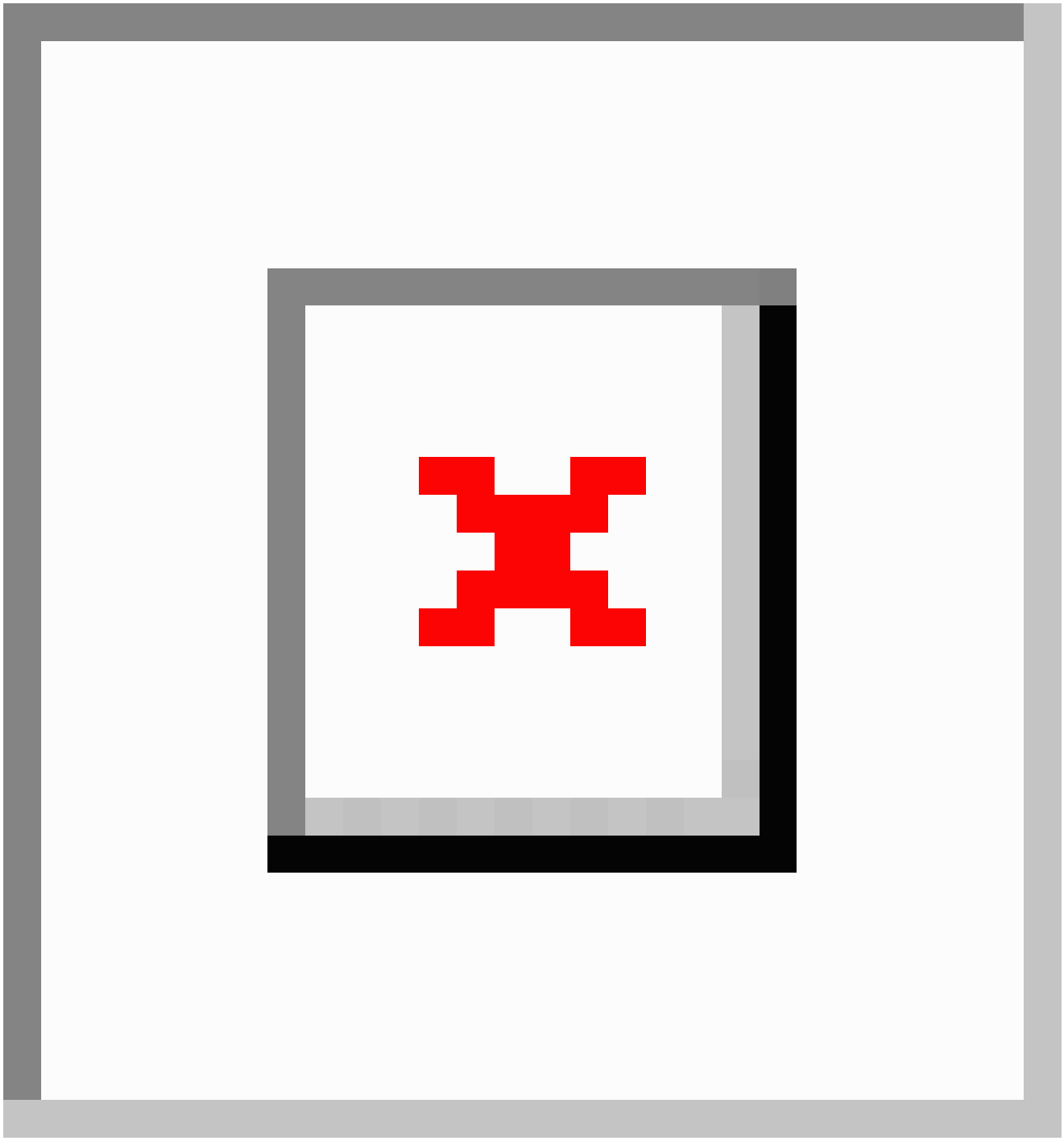
Differences Between Devices

To determine whether there were any reliable differences between devices in their subjective usability, a one-way analysis of variance was conducted with the SUS as the dependent variable and medical device as the independent variable. This analysis was only done for those devices that had more than 10 responses, which included the following: thermometer, blood

pressure cuff, inhaler, pregnancy test kit, syringe, blood glucose meter, and epinephrine injector (EpiPen).

Figure 1 shows the average SUS scores by devices for those devices that had more than 10 responses. As seen in this figure, the thermometer had the highest score and the EpiPen had the lowest. The effect of device on SUS scores was significant, $F_{6,413}$=7.27, P<.001, η^2=.096, and a Tukey post hoc analysis found that the thermometer usability was significantly higher than the blood pressure cuff (P=.014), EpiPen (P=.018), pregnancy test kit (P=.001), and syringe (P=.02).

Figure 1. Mean System Usability Scale scores by device. Error bars represent the 95% CI.



Discussion

Principal Findings

In this study, we examined the usability of a number of home health care devices using a field-based retrospective methodology and the SUS. The study was designed to determine the usability characteristics of these medical devices in realistic settings, and set the stage for much larger data-collection efforts using this method in the future, to much more fully characterize the usability of home health care devices. This study yielded several important findings. First, even with this young, healthy, and well-educated sample population, a wide variety of home health care devices was used. Second, the usability of these rated devices covered a fairly wide range, and third, there were statistically significant differences (see the “Results” section)

in the subjective usability ratings given to these different devices. These findings suggest that this method could be used to great effect to more fully characterize a broader range of home health care devices.

Because medical devices in the home are often used in critical life and death situations and can also be an important part of maintaining a healthy life, it is essential to understand how usable these home health care devices are. If a user cannot successfully use their complex new television remote control, then the result is simply an inability to watch television. If a user fails to successfully and correctly use a home health care device, the impact could be significantly greater, up to and including death. This study used a convenience sample of young, well-educated users to make these evaluations. From the results obtained, it is reasonable to be concerned that people who are ill and using more complex devices will have similar or (likely) worse experiences with their home health devices.

Many home health care devices are, in large part, another consumer item. They are widely available to the general public and are sold in both traditional brick-and-mortar retail outlets and through general merchandize online outlets such as Amazon.com, Inc. Because many of these devices are no longer the sole purview of specialized medical device retailers, it seems likely that consumers may view these devices as another commodity and will make assessments of the usability and utility of home health care devices in the same way that they make assessments of other consumer goods. The question of how the usability of these home health care devices compares to other common devices used by the general public is instructive because one would expect (and hope) that home health care devices would have higher ease of use characteristics, given the importance of their function. Figure 2 shows how these home health care devices compare to 14 other kinds of commonly used software programs and devices that were described by Kortum and Bangor [18]. Remarkably, the rated medical devices are some of the most unusable. As can be seen, they occupy 5 of the 7 lowest scores, when compared with these common devices used by the general public. Only the inhaler and the thermometer score in the middle of the pack. Of particular note is the rating given to the EpiPen. This is a device that must be used correctly, at a time and place not of the user's choosing, under conditions that can only be described as exceedingly stressful. Failure to use the device correctly within a very small time frame can result in death. There is no time to consult the instruction manual and no time to call for technical support. And yet, even with this mission criticality, the device was rated very low in its subjective usability.

If we plot these devices on the Usability Acceptability Scale [24], it can be seen (Figure 3) that over half of the rated devices are in the "marginal" range, with the remaining ones being judged as "acceptable." Clearly, no medical device should be in the "marginal" or "unacceptable" ranges, particularly those that have life-or-death consequences.

Figure 2. Comparison of usability ratings for the home health care devices in this study (black bars) and 14 common products described by Kortum and Bangor [18] (gray bars).

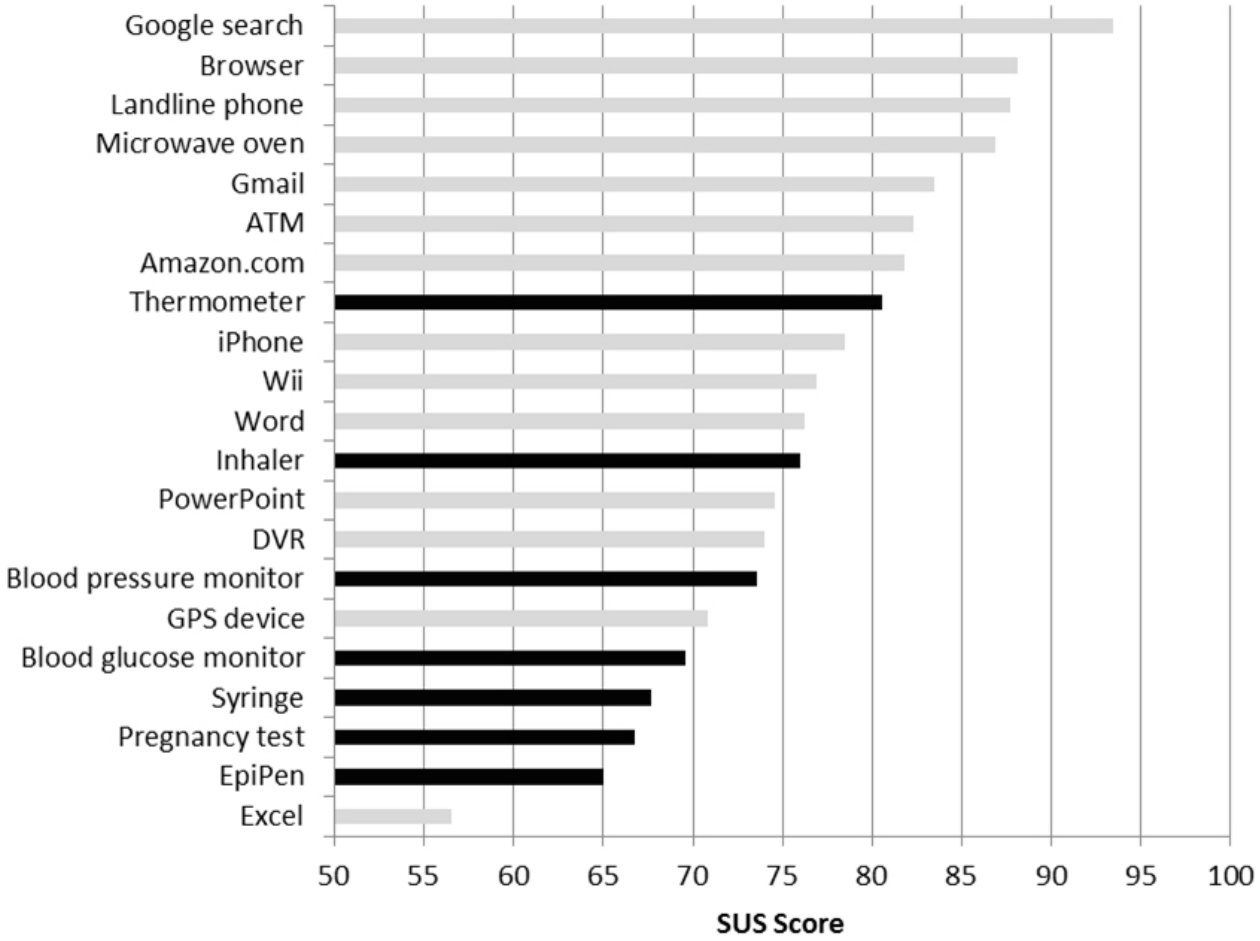
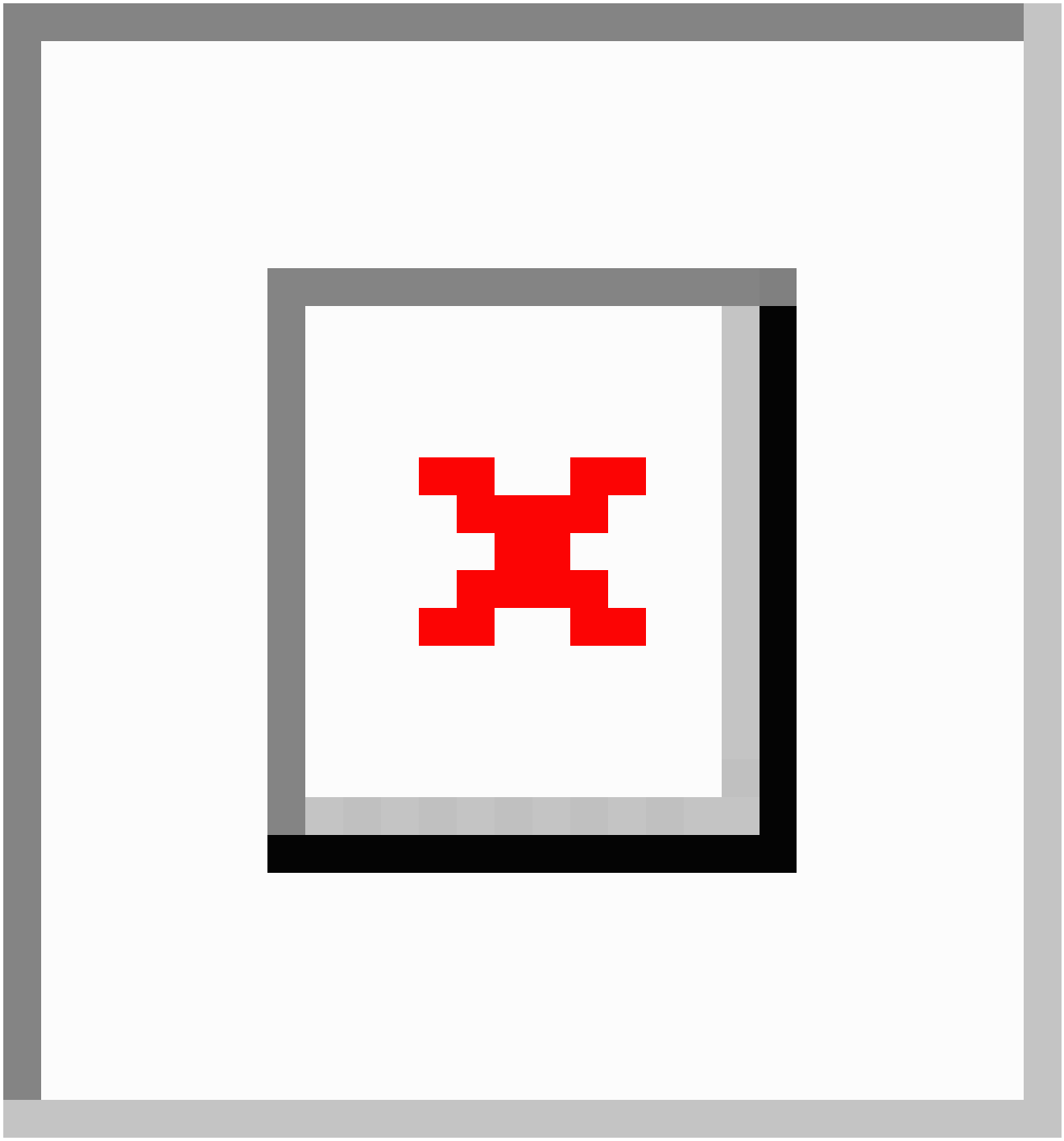


Figure 3. Usability ratings of the home health care devices plotted on the acceptability scale [24]. Scores below 50 are judged to be unacceptable.



Future Directions

The data show that there is wide variability in the usability of different home health devices, even among a relatively young, healthy population of users. Further research needs to be undertaken to explore how usability ratings may differ by demographic variables such as health status, age, socioeconomic status, and education levels. With this expansion in demographics, the classes of devices that users can rate will also need to be expanded to accommodate home health devices that are used more in chronic care (eg, dialysis machines, lift equipment). This expanded data would also allow for the construction of more relevant comparisons with the usability of other home health care devices, rather than just with common consumer devices, as shown in Figure 2. This would more accurately reflect the frequency (eg, regular use of an inhaler and infrequent use of an EpiPen) and nature of the interactions (eg, critical versus noncritical) that occur with home health care devices in the field. Further research should also be conducted to determine whether the adjective rating scales found with general consumer products (Figure 3) are still appropriate for home health care devices.

This kind of future research would set the stage for communicating with physicians, hospitals, and patients about the specific kinds of home health care devices which have sufficiently poor usability that their use, as is typically prescribed, might represent a risk to the health of the patients using them. From that information, further work can be

undertaken to determine what could be done to mitigate or eliminate these risks.

These steps might include working directly with manufacturers and physicians to identify methods of directly providing information about the need for increased patient contact and training when certain classes of devices are prescribed. For example, it could be that when physicians prescribe the use of a home health care device, a "usability risk database" is referenced, which alerts the physician that he or she will need to follow-up more frequently with patients using these devices. Manufacturers could communicate with the users of devices through their instruction books or warning labels to alert users of low-usability/high-risk devices that "this piece of equipment must involve extensive training before use in a home health environment." In the longer term, a dissemination mechanism, such as a website, could be constructed such that consumer groups, physicians, manufacturers, patients, and caregivers could search for usability information for specific types of home health care devices (akin to Consumer Reports). This website could be linked to major sources of medical information such as WebMD or Wikipedia, making it readily and easily available when patients and caregivers are making health decisions.

All of these dissemination mechanisms would have the sole goal of educating critical personnel in the care chain (physicians, nurses, patients, and patient caregivers) about the usability of a wide range of home health care devices. Consumers have ready access to this kind of information for all manners of other consumer goods, but there is a gap when it comes to many home medical devices. The method of remotely collecting usability data described in this paper would allow for the creation of these kinds of medical device usability information databases. These databases would, in turn, provide valuable information on the usability of devices throughout their life cycle.

Although there are many benefits to be derived from a larger scale collection of subjective usability data for home health care devices, interpretation of the data must be done carefully. The correlational relationship between task performance and subjective usability assessment is not perfect [23]. There may be cases where devices have acceptable task performance in the field, but are judged poorly with subjective usability measures. In this case, the need for further attention to the device would be captured and the benefits of additional design work could be measured against the time and cost of modifying a device that has otherwise sufficient performance properties. More concerning would be those devices that receive acceptable-to-high subjective usability scores, but have poor performance characteristics. In this case, the need for further attention to these devices might not be noted, because performance data are not specifically captured in this remote protocol. Further research should be conducted to determine the relationship between subjective and objective usability measures for home health care devices and if there are methods to accurately capture device performance elements from the questionnaire format.

Conclusions

Understanding the usability of home health care devices is important as more health care is pushed into the home. Patients, who used to be cared for primarily in hospitals or long-term care facilities, are now routinely sent home with a myriad of medical devices to manage and treat their conditions. With a sufficiently expanded data-collection effort, the kind of usability data described here could be used to impact not only the design and development of future devices, but also could be used immediately to help physicians and patients alike make better, more informed decisions when prescribing or choosing home health care equipment.

As always, the more information a physician and patient can share about the patient's care, the better that care can be. The continuing transition from hospital care to home health care means that the usability of devices now used in the home needs to more fully understood, and this information needs to be shared, so that care can be delivered in the safest, most effective possible way.

Conflicts of Interest

None declared.

References

1. Marcus A. User interface design's return on investment: Examples and statistics. In: Bias RG, Mayhew DJ, editors. Cost Justifying Usability: An Update for the Internet Age. Amsterdam, The Netherlands: Elsevier; 2005:17-39.
2. Kohn LT, Corrigan JM, Donaldson MS. To Err Is Human: Building a Safer Health System. Washington, DC: National Academies Press; 2000:1-311.
3. James JT. A new, evidence-based estimate of patient harms associated with hospital care. J Patient Saf 2013 Sep;9(3):122-128. [doi: 10.1097/PTS.0b013e3182948a69] [Medline: 23860193]
4. National Association for Home Care and Hospice. Basic Statistics About Home Care. Washington, DC: National Association for Home Care and Hospice; 2010. URL: http://www.nahc.org/assets/1/7/10HC_Stats.pdf [accessed 2015-04-16] [WebCite Cache ID 6XrPTx9HN]
5. Fisk AD, Rogers WA, Charness N, Czaja SJ, Sharit J. Designing for Older Adults: Principles and Creative Human Factors Approaches. London, UK: CRC press; 2009.
6. United States Food and Drug Administration. Medical Device Home Use Initiative. Silver Spring, MD: United States Food and Drug Administration; 2010. URL: http://www.fda.gov/downloads/UCM209056.pdf [accessed 2015-05-31] [WebCite Cache ID 6YxzCfBN0]

7. Fairbanks RJ, Shah MN, Caplan S, Marks A, Bishop P. Defibrillator usability study among paramedics. Proc Hum Fact Ergon Soc Annu Meet 2004;48(15):1768-1772. [doi: 10.1177/154193120404801530]
8. Gundry JW, Comess KA, DeRook FA, Jorgenson D, Bardy GH. Comparison of naive sixth-grade children with trained professionals in the use of an automated external defibrillator. Circulation 1999;100(16):1703-1707. [doi: 10.1161/01.CIR.100.16.1703]
9. Wallace LS, Zite NB, Homewood VJ. Making sense of home pregnancy test instructions. J Womens Health (Larchmt) 2009;18(3):363-368. [doi: 10.1089/jwh.2008.0985] [Medline: 19243273]
10. Surabattula D, Harvey CM, Aghazadeh F, Rood J, Darisipudi A. Usability of home cholesterol test kits and how their results impact patients' decisions. Int J Ind Ergon 2009;39(1):167-173. [doi: 10.1016/j.ergon.2008.07.006]
11. Santos L, Damiani M, Vaughn-Cooke M. Designing usable glucometers to support patient adherence. 2013 Presented at: Proceedings of the Human Factors and Ergonomics Society Healthcare Symposium; Mar 10â€"13, 2013; Baltimore, MD p. 04-14 URL: http://www.webcitation.org
12. American National Standards Institute, Association for the Advancement of Medical Instrumentation. Human Factors Engineering: Design of Medical Devices (ANSI/AAMI HE75:2009). Washington, DC: American National Standards Institute; 2009:1-26.
13. van Dulmen S, Sluijs E, van Dijk L, de Ridder D, Heerdink R, Bensing J. Patient adherence to medical treatment: A review of reviews. BMC Health Serv Res 2007 Apr 17;7:55 [FREE Full text] [doi: 10.1186/1472-6963-7-55] [Medline: 17439645]
14. Holden RJ, Karsh BT. The technology acceptance model: Its past and its future in health care. J Biomed Inform 2010;43(1):159-172 [FREE Full text] [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
15. ECRI Institute. Top 10 technology hazards for 2012. The risks that should be at the top of your prevention list. Health Devices 2011 Nov;40(11):358-373. [Medline: 23444580]
16. International Organization for Standardization. Ergonomic requirements for office work with visual display terminals (VDTs): ISO 9241-11:1998. Geneva, Switzerland: The International Organization for Standardization; 1998.
17. Schmettow M, Vos W, Schraagen JM. With how many users should you test a medical infusion pump? Sampling strategies for usability tests on high-risk systems. J Biomed Inform 2013;46(4):626-641 [FREE Full text] [doi: 10.1016/j.jbi.2013.04.007] [Medline: 23688827]
18. Kortum PT, Bangor A. Usability ratings for everyday products measured with the System Usability Scale. Int J Hum Comput Interact 2013 Jan;29(2):67-76. [doi: 10.1080/10447318.2012.681221]
19. Kortum P, Peres SC. The relationship between system effectiveness and subjective usability scores using the System Usability Scale. Int J Hum Comput Interact 2014 Jul;30(7):575-584. [doi: 10.1080/10447318.2014.904177]
20. Brooke J. SUS: a 'quick and dirty' usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland IL, editors. Usability Evaluation in Industry. London, UK: Taylor and Francis; Jun 11, 1996:189-194.
21. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. Int J Hum Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
22. Kortum P, Acymyan CZ. How low can you go? Is the System Usability Scale range restricted? J Usability Stud 2013 Nov;9(1):14-24.
23. Sauro J. A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices. Denver, CO: Measuring Usability LLC; 2011.
24. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: Adding an adjective rating scale. J Usability Stud 2009;4(3):114-123.
25. Story MF. Medical devices in home health care. In: Olson S, editor. The Role of Human Factors in Home Health Care: Workshop Summary. Washington, DC: National Academies Press; 2010.

Abbreviations

AAMI: Association for the Advancement of Medical Instrumentation
AED: automatic external defibrillator
EEG: electroencephalography
FDA: United States Food and Drug Administration
IOM: Institute of Medicine
SUS: System Usability Scale

Edited by G Eysenbach; submitted 22.04.15; peer-reviewed by J Brooke; comments to author 19.05.15; revised version received 21.05.15; accepted 21.05.15; published 05.06.15.

Please cite as:
Kortum P, Peres SC
Evaluation of Home Health Care Devices: Remote Usability Assessment
JMIR Human Factors 2015;2(1):e10
URL: http://humanfactors.jmir.org/2015/1/e10/
doi:10.2196/humanfactors.4570
PMID:27025664

Original Paper

Usability and Acceptability of ASSESS MS: Assessment of Motor Dysfunction in Multiple Sclerosis Using Depth-Sensing Computer Vision

Cecily Morrison[1*], PhD; Marcus D'Souza[2*], MD; Kit Huckvale[1,3*], MB ChB, MSc; Jonas F Dorn[4], PhD; Jessica Burggraaff[5], MD; Christian Philipp Kamm[6], MD; Saskia Marie Steinheimer[6], MD; Peter Kontschieder[7], PhD; Antonio Criminisi[7], PhD; Bernard Uitdehaag[5], MD; Frank Dahlke[4], MD, PhD; Ludwig Kappos[2], MD; Abigail Sellen[1], PhD

[1]Microsoft Research, Human Experience & Design, Cambridge, United Kingdom

[2]Neurology, Department of Medicine, Clinical Research, Biomedicine and Biomedical Engineering, University Hospital Basel, Basel, Switzerland

[3]Global eHealth Unit, Imperial College London, London, United Kingdom

[4]Novartis Pharma AG, Basel, Switzerland

[5]VU University Medical Center, Amsterdam, Netherlands

[6]Department of Neurology, University Hospital Bern, Bern, Switzerland

[7]Microsoft Research, Machine Learning and Perception, Cambridge, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Cecily Morrison, PhD
Microsoft Research
Human Experience & Design
21 Station Road
Cambridge,
United Kingdom
Phone: 44 1223 479951
Fax: 44 1223 479700
Email: cecilym@microsoft.com

Abstract

Background: Sensor-based recordings of human movements are becoming increasingly important for the assessment of motor symptoms in neurological disorders beyond rehabilitative purposes. ASSESS MS is a movement recording and analysis system being developed to automate the classification of motor dysfunction in patients with multiple sclerosis (MS) using depth-sensing computer vision. It aims to provide a more consistent and finer-grained measurement of motor dysfunction than currently possible.

Objective: To test the usability and acceptability of ASSESS MS with health professionals and patients with MS.

Methods: A prospective, mixed-methods study was carried out at 3 centers. After a 1-hour training session, a convenience sample of 12 health professionals (6 neurologists and 6 nurses) used ASSESS MS to capture recordings of standardized movements performed by 51 volunteer patients. Metrics for effectiveness, efficiency, and acceptability were defined and used to analyze data captured by ASSESS MS, video recordings of each examination, feedback questionnaires, and follow-up interviews.

Results: All health professionals were able to complete recordings using ASSESS MS, achieving high levels of standardization on 3 of 4 metrics (movement performance, lateral positioning, and clear camera view but not distance positioning). Results were unaffected by patients' level of physical or cognitive disability. ASSESS MS was perceived as easy to use by both patients and health professionals with high scores on the Likert-scale questions and positive interview commentary. ASSESS MS was highly acceptable to patients on all dimensions considered, including attitudes to future use, interaction (with health professionals), and overall perceptions of ASSESS MS. Health professionals also accepted ASSESS MS, but with greater ambivalence arising from the need to alter patient interaction styles. There was little variation in results across participating centers, and no differences between neurologists and nurses.

Conclusions: In typical clinical settings, ASSESS MS is usable and acceptable to both patients and health professionals, generating data of a quality suitable for clinical analysis. An iterative design process appears to have been successful in accounting for factors that permit ASSESS MS to be used by a range of health professionals in new settings with minimal training. The study

shows the potential of shifting ubiquitous sensing technologies from research into the clinic through a design approach that gives appropriate attention to the clinic environment.

(JMIR Human Factors 2015;2(1):e11) doi:10.2196/humanfactors.4129

KEYWORDS

depth-sensing computer vision; information interfaces and presentation; Kinect; motor skills; multiple sclerosis; rehabilitation

Introduction

Overview

The recent development of robust depth-sensing cameras for practically tracking human motion is being rapidly exploited for the assessment and rehabilitation of motor dysfunction. Depth-sensing cameras, such as Microsoft Kinect (Microsoft, Redmond, WA, USA), capture video images in which each pixel has a three-dimensional position. Processed by advanced computer vision and machine-learning algorithms, depth videos enable the quantification of human movement without the need for marker-based motion capture or gait analysis systems, which are both expensive and cumbersome [1]. A particular advantage is that nothing needs to be attached to the patient.

Depth sensing has been used to build a range of health care applications, including touchless interaction during surgical image navigation[2], movement rehabilitation for children with cerebral palsy [3], and improving cognitive performance in elderly people [4]. Within the domain of multiple sclerosis (MS), depth sensing has been used to improve posture during exercise for patients with MS [5] and incorporated into an exercise game (telerehabilitation system) to encourage balance and sensory integration [6]. Virtual reality games have also been used to motivate motor rehabilitation exercises [7].

The use of depth sensing has only recently been extended to the assessment of motor dysfunction. Although sensing technology has the potential to increase the reliability and validity of assessment compared with human observers [8], achieving the high levels of system accuracy required for diagnostic purposes has proven challenging. Research has mainly focused on the validation of Kinect against other objective measurement systems [9] and its ability to provide accurate measures for particular conditions [10,11]. Systems that successfully provide clinical assessment of motor ability with Kinect remain very much “in progress” [12].

Background

ASSESS MS is being developed to support the assessment of motor dysfunction in people with MS. As a chronic inflammatory disease of the central nervous system, MS causes a variety of symptoms, either in combination or alone. These include numbness, reduction in motor strength, and cerebellar dysfunctions as well as cognitive decline. The disease course is most frequently characterized by relapses in which the affected person experiences neurological symptoms followed by extended periods of remission in which symptoms may improve. Over time, the disease can enter into a progressive phase in which a steady deterioration occurs. Approximately 15% of MS patients experience ongoing deterioration from disease onset [13].

The unpredictability of the disease course makes the ability to track MS particularly useful. The condition is currently assessed with a standardized rating instrument based on clinical examination, the Expanded Disability Status Scale (EDSS) [14]. Patients are asked to perform a range of functional exercises, including stretching out 1 arm to the side and then touching the nose (finger-nose test) or walking on a pretend tightrope (tightrope walking). These exercises are summarized into functional system scores and, together with the ability to walk, are scored on an ordinal rating scale, from EDSS 0 to EDSS 10. Examinations are usually performed on a yearly basis. Although the EDSS is a widely used and accepted outcome measure, it suffers from low intrarater and inter-rater reliability making disease tracking difficult [15]. The expertise required also makes it infeasible for health professionals other than neurologists to perform the examination.

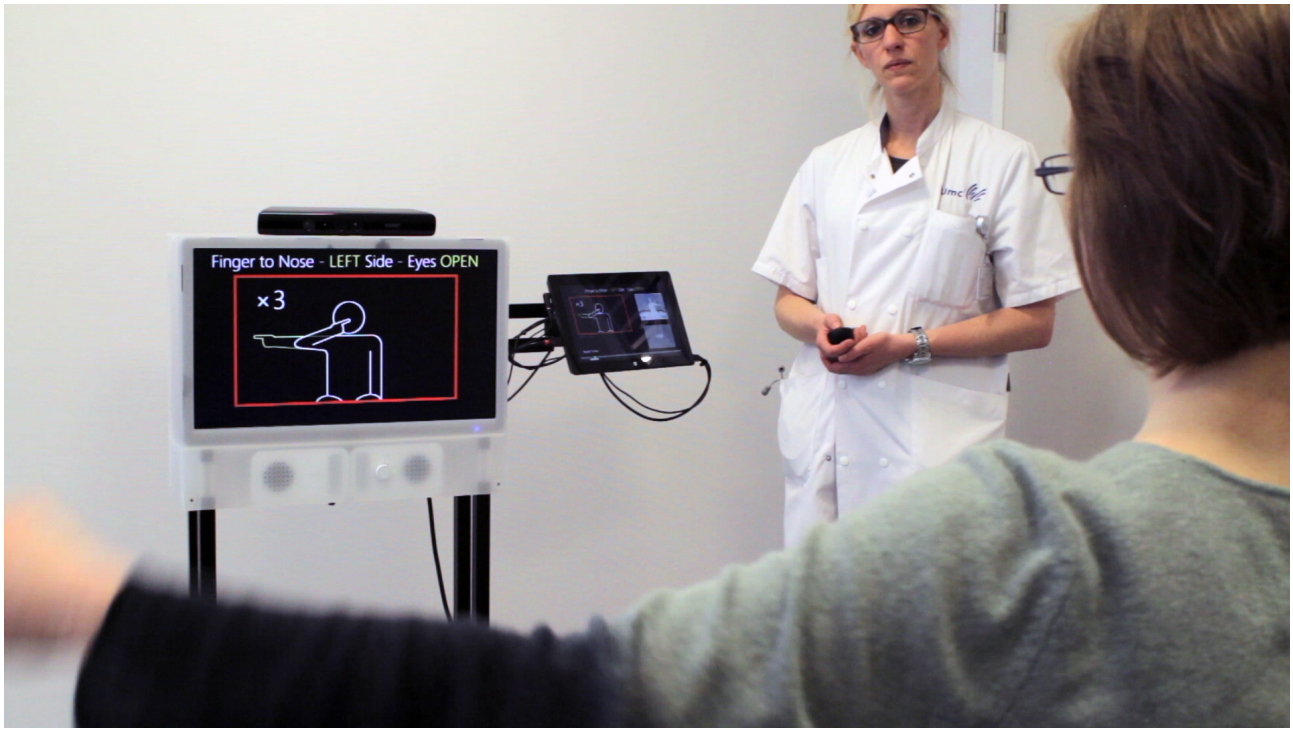
ASSESS MS aims to address this problem by quantifying changes in motor dysfunction more consistently and with finer granularity than currently possible. Shown in Figure 1, ASSESS MS captures depth videos of assessment movements with Kinect, which were performed by patients in a clinical setting with the support of the health professional. These are then processed to classify the severity of motor dysfunction. The level of accuracy needed for clinical assessment requires specific attention to the quality of the depth videos captured. Specifically, a high level of standardization is required.

The inherent unpredictability of hospital environments and the need for highly standardized data make the step from validating depth-sensing measures in the laboratory to creating a workable system in a clinical setting nontrivial [2]. Unintended variability, whether from inconsistent movement performance or poor image quality due to variable positioning or unexpected objects in the background, decreases the likelihood that the vision algorithms will be able to highlight variability that arises from disease. At present, there are no studies that show that this step is feasible for the clinical assessment of motor dysfunction with depth-sensing computer vision.

A key element of ASSESS MS is the design of both the physical device and the software application to support high-quality data capture in the clinical environment. The study presented here is a mixed-methods empirical evaluation of the usability and acceptability of these aspects of ASSESS MS. It aims to answer the following questions:

- Is ASSESS MS usable by health professionals?
- Is ASSESS MS acceptable to patients and health professionals?
- Are there any differences between neurologists and nurses in any of the metrics captured?

Figure 1. ASSESS MS being used to record a finger-nose test while the health professional monitors.



Methods

System Development

ASSESS MS was developed by a multidisciplinary team of researchers in human-computer interaction, machine learning, and health professionals caring for patients with MS. A 4-stage, iterative development process was undertaken to develop the physical device and software application used to support the recording and analysis of movements. Its design drew on multidisciplinary team meetings, design activities, observation of current clinical assessment practice, and user testing with healthy volunteers and patients. We first present the problems noted in the early stages of the project and the resulting design requirements before presenting the final system design. Algorithmic development is reported elsewhere [16].

Design Requirements

Observation of clinical routines identified several issues that needed to be addressed. We observed that a neurologist could instruct the same movement in different ways. The finger-nose test, for example, might begin by stretching the hand out to the side and in other cases stretching the hand to the front before touching the nose. Clinicians were also observed to adapt the instructions given to a patient according to their abilities. For example, while able patients might be asked to bicycle their legs to assess strength, patients with a degree of disability would be asked to push their leg against the clinician's hand.

This kind of variation, although of no consequence to a neurologist, is problematic for machine-learning algorithms, which statistically evaluate patients against "known" characteristics derived from training examples. If a large amount of variation in movement performance arises from factors other than those relating to disease state, classification ability inevitably decreases. The findings emphasized the need to not only provide cues to standardize the movement, but also offer discretion for health professionals to omit movements if necessary, as well as repeat them if performed incorrectly.

We also noted that current examinations are an embodied interaction between clinician and patient. It was not uncommon for a clinician to stand in front of a patient and demonstrate the movement to be performed. A clinician might also touch the patient to indicate how to do a movement or which side to use. Most importantly, the clinician may have to stand next to a patient due to safety reasons. Many of these typical interactions had the potential to disturb the image captured by ASSESS MS, either by blocking the camera view or creating a challenge to distinguish between patient and doctor. At the same time, appropriate patient-health professional interactions are important to ensure patients feel safe and cared for.

Not least, the clinical examination is a mobile affair. Patients can move large distances while performing a movement, such as hopping on 1 foot. Patients may perform small movements, adjusting their sitting position, which can lead to limbs being out of the camera view. In smaller rooms, furniture needs to be adjusted and the camera moved multiple times between different types of tests (eg, sitting and standing tests), which can change the camera view and the placement of the patient in that view. With all of this movement, it was necessary to achieve as standard a lateral and depth positioning as possible, to facilitate the preprocessing of the videos.

This initial work suggested that ASSESS MS needed to support the following aspects of capture:

- Standardized movement instruction;
- Flexible interface for the health professional to record movements;
- Facilitated patient-health professional interaction with maintained image quality; and

- Precise positioning of the patient.

ASSESS MS Description

ASSESS MS, shown in Figure 2, has a 53.3-cm (21-in) patient-facing screen used to instruct patients in the assessment movements. A smaller tablet computer with touch-screen capability is mounted on a mobile arm at the back of the unit. This interface is used by the health professional to position the patient, select the assessment movements to be performed, and complete the recordings. A remote control enables the health professional to move freely around the room to support the patient as needed. The screens sit in an ergonomic box on wheeled legs for ease of maneuvering with the Kinect mounted on top.

The health professional interface provides a number of navigational options. The health professional can play a movement instruction video, or begin a test. Arrows at the top of the interface enable the health professional to skip movements (eg, finger-nose test) or variations of movements (eg, left side, eyes open). Movements can be repeated by skipping backward. Each page contains a button, which enables the beginning of a test, recording of a movement, or stopping of a recording. A navigational bar at the bottom shows visually which movements have been captured and which skipped. These are all shown in the top image of Figure 3.

Precise support for positioning is provided through an easily maneuverable device used in conjunction with the "positioning" feature, as shown in the middle image of Figure 3. This screen provides a view of the depth image stream with a center crossbar to which the patient should be aligned. It is available before the sitting and standing components in the test as a full-screen feature and in a persistent window in the upper-right-hand corner throughout. The distance of the person from the camera is indicated below the image. It was intended to reduce variability in positioning.

Movement instructional videos are provided to standardize movement performance. They guide the patient, as well as the health professional, about exactly how to perform the requested movement. They consist of simple line drawing animations accompanied by verbal descriptions localized into 3 languages. The design of the animations was based on the psychology literature on movement learning, which emphasized simplicity of representation [17] and the importance of drawing attention to the most distal point of movement, for example, the hand when moving the arm [18].

A number of approaches were taken to support the patient-health professional interaction in light of the instructional videos that change the nature of this interaction. First, the placement of the health professional interface is intended to encourage health professionals to stand to the side or back of the device to avoid blocking the camera view. Second, automatic recording of movements was not used to enable appropriate pauses in the examination to facilitate interaction.

Figure 2. Elements of ASSESS MS, including instructional system, Kinect depth-sensing camera, health professional interface, remote control, and ergonomic box.

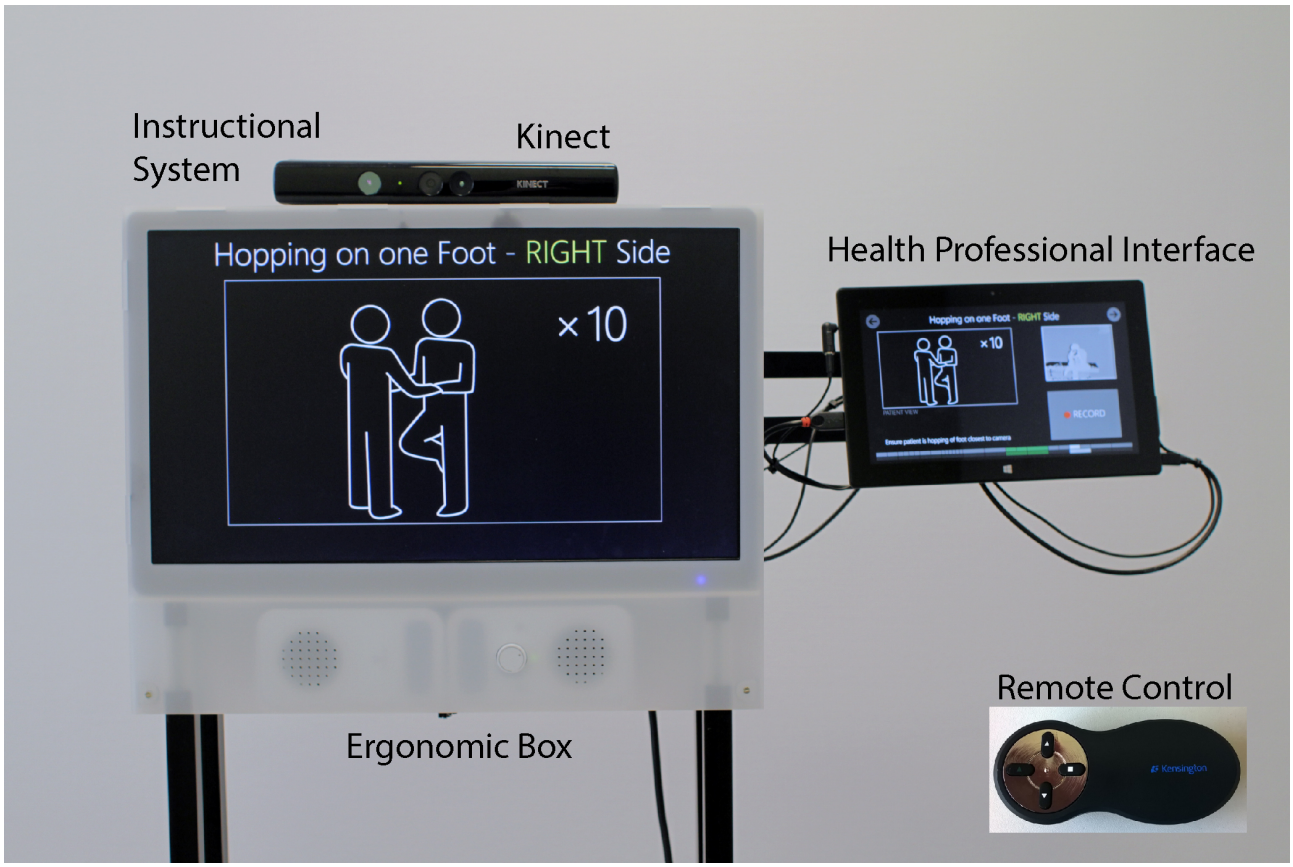
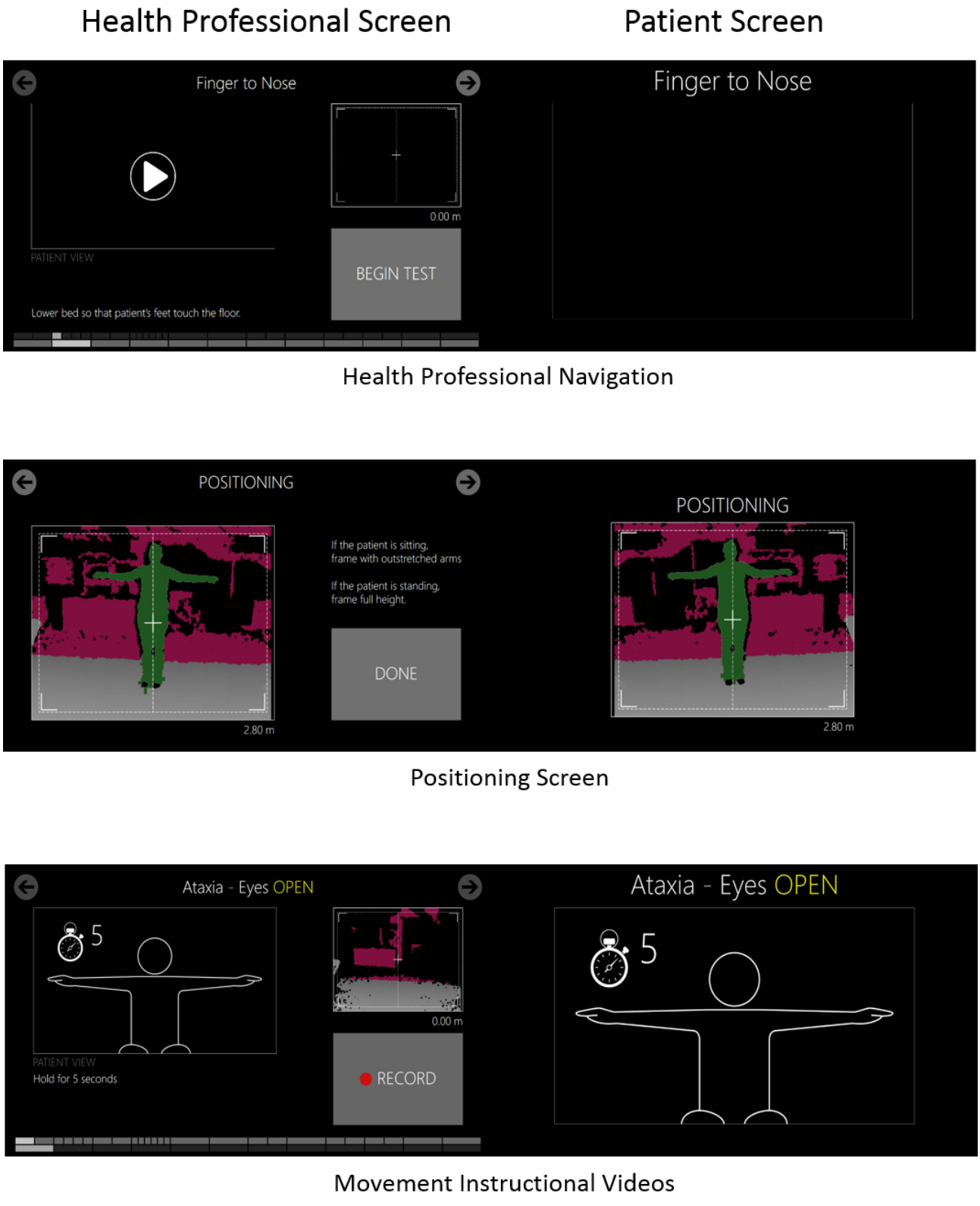


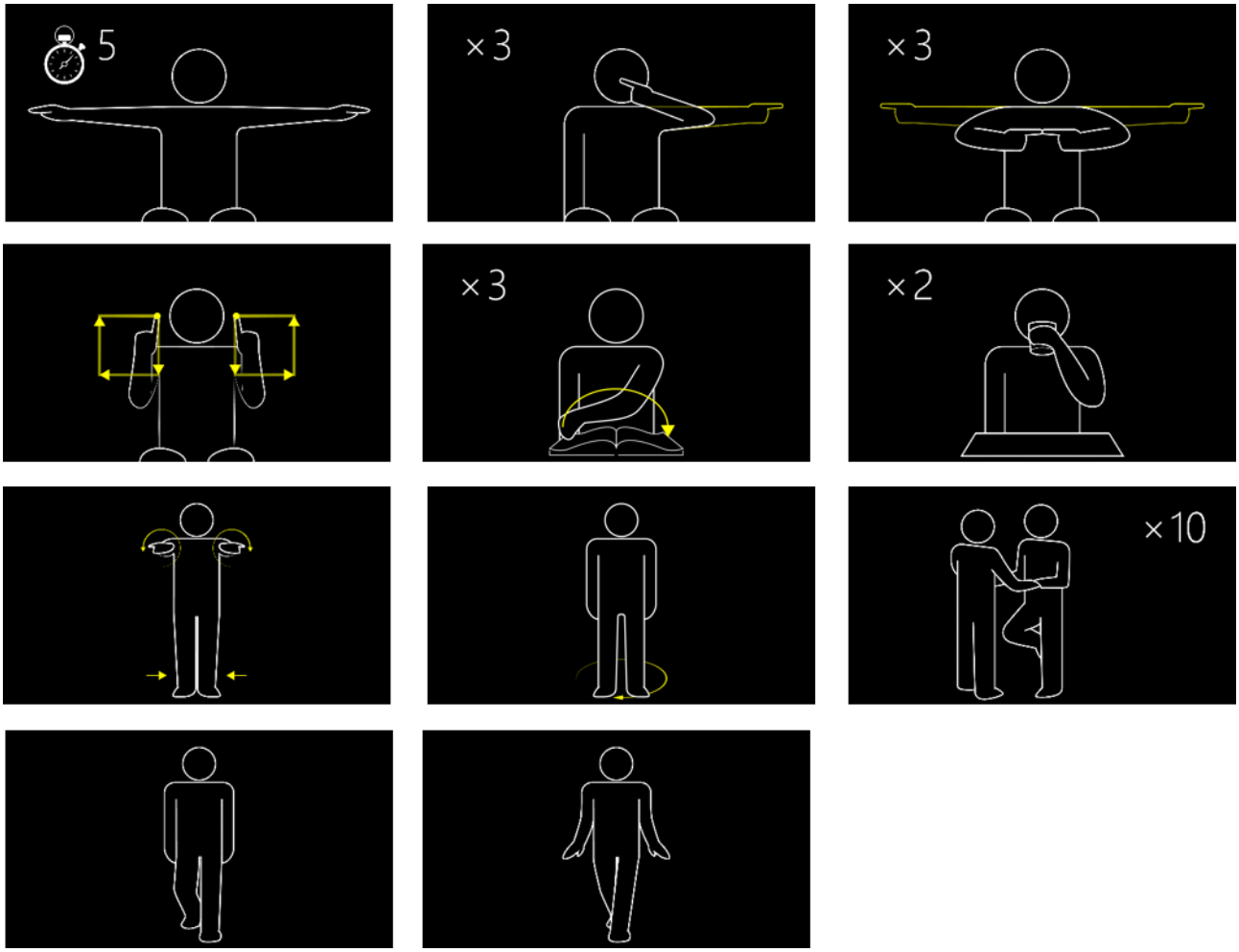
Figure 3. Screenshots of interface: health professional screen on the left and patient screen on the right. Top, navigational elements; middle, positioning feature; bottom, movement instructional videos.



Movement Protocol

The movement protocol contained 11 movements depicted in Figure 4. Of these 11 movements, 6 were chosen from the EDSS examination to cover the function of the upper and lower extremities and the trunk. Two activities of daily living movements, drinking from a cup and turning pages in a book, were also included. In addition, 3 new movements were defined, which included finger-finger test, drawing squares, and rotating on the spot. These were created to capture the upper and lower extremity functions in a potentially more camera-friendly way.

Figure 4. The 11 assessment movements in the ASSESS MS movement protocol: truncal ataxia, finger-nose test, finger-finger test, drawing squares, turning pages, drinking from cup, Romberg test, turning on the spot, hopping on 1 foot, normal walking, and tightrope walking.



Study Design

Health professionals, previously unfamiliar with ASSESS MS, were asked to use it to examine 4 MS patients following 1 hour of training. The training covered the importance of standardized movement performance, the movement protocol, and the features to promote image quality. A “cheat” sheet was given to all of the health professionals with the details of the movement protocol for ease of reference (see Multimedia Appendix 1). The set of examinations done by a given health professional took place within a week, and most on a single day.

Patients were included if theyhad a diagnosis of MS and an EDSS score between 0 and 7. After giving informed, written consent, patients were randomly assigned to a health professional. To minimize inconvenience to patients, invitations to act as participants were extended to individuals already attending routine clinic appointments during the study period. Some patients were given an examination with the same movement protocol [16], within a parallel ongoing trial (n=10), but none within the previous 3 months. Ethical approval was obtained in all 3 hospitals.

Outcomes

Usability has been defined by the International Organization for Standardization in ISO 9241-11 [19] as “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.” In this case, we wanted to achieve the following 4 goals:

1. Complete a full protocol of recordings, repeating, or skipping movements as necessary.
2. Obtain standardized movement performance from patients.
3. Position ASSESS MS adequately for quality data capture.
4. Ensure the camera view is not blocked by the health professional.

The *context of use* is defined as a clinical setting.

We focused on 2 types of *users* (neurologists and nurses). While neurologists are the group that currently performs neurological examinations in clinical trials, it would be more cost effective if semiautomated examinations could be completed by other health professionals. These might include nurses, study nurses, or paramedic staff. As most will be nurses, we refer to them hereafter as such. Recognizing this range of potential users, ASSESS MS was deliberately designed for nonexperts with minimal training.

The way that the ISO definition of usability is tested depends very much on the technology. We have articulated specific metrics for effectiveness, efficiency, and satisfaction as detailed in Table 1 that match the goals of the system and context of use specified earlier.

Acceptability is part of the ISO definition of usability under the term "satisfaction." However, we take a view of acceptability more consistent with studies in the clinical domain. We assume that system acceptance is the trade-off between the benefit provided and any discomfort that arises, rather than some inherent good feeling that is gained. It is measured through willingness for future use, impact on patient-health professional interaction, and general perceptions of the technology as shown in Table 1.

Table 1. Description of metrics used for each trait and the data source drawn upon.

Trait[a,b]	Metric	Data source
Effectiveness	Task completion	Video
	Movement performance standardization	Depth recordings
	Adequate positioning	Depth recordings
	Clear camera view	Depth recordings
	Clarity of task	Questionnaire (Q1 and Q3)
Efficiency	Time to completion	Video
	Perceived efficiency	Interviews
Satisfaction	Covered by acceptability	
Future use	Willingness to use again	Questionnaire (Q2)
Interaction	Attitudes toward human interactions	Questionnaire (Q4 and Q5)
	Articulated changes to patient-health professional interactions	Interviews and video
Perceptions	Description of recording tool and associated issues	Patient questionnaire (Q6-Q9) and interviews

[a]Effectiveness, efficiency, and satisfaction constitute "usability."

[b]Future use, interaction, and perceptions constitute "satisfaction."

Data Collection

In addition to the depth recordings taken as part of the ASSESS MS examination, all assessments were video recorded with a separate digital video camera that captured the way health professionals and patients interacted in the examination room.

A 9-item questionnaire was given to patients at the end of the examination. It contained 5 Likert-scale-type questions and 4 free response questions. The Likert-scale questions were developed by researchers in human-computer interaction (CM and AS) based on 2 key usability constructs: ease of use and impact on human interaction. Q1 and Q3, respectively, queried ease of use of the whole system and its instructional aspects alone. Q5 and Q4 queried the same for impact on human interaction. Q2 focused on the acceptability of ASSESS MS as determined by willingness for future use. Positively and negatively framed questions were counterbalanced.

Four open-end questions were asked to enable an opportunity for patients to give a more extensive commentary on their experience. The first asked patients to characterize ASSESS MS for another patient in 3 words. The second and third focused on the most and least helpful aspects of ASSESS MS. The fourth and final question provided space for any further comments. Questionnaires were translated into German and Dutch by the respective clinical teams, and wording for this translation was agreed internally. Local piloting with patients was carried out in 1 clinic in which multiple clinician input could not be obtained.

Health professionals were given a similar 5-item Likert-scale-type questionnaire after each examination. The questions were intended to be equivalent to the patient questionnaire for comparative purposes. Following the completion of questionnaire by all patients, professionals also took part in a 15-minute debriefing interview. The questions were similar to the free response questions given to the patients, but done in an interview form to enable more extensive discussion (see Multimedia Appendix 2).

To assess whether the results applied to a wide spectrum of patients, the EDSS and the Symbol Digit Modalities Test (SDMT) scores were recorded for each patient. Calculated from a detailed neurological examination, the EDSS was used to assess physical disability. The scale includes 20 half steps, ranging from 0 (normal) to 10 (death due to MS) [14]. The SDMT was used as a measure of cognitive ability. It examines information processing speed, visual working memory, and concentration by primarily assessing complex and visual scanning and tracking. It ranges from 0 (no correct answers) to 110 (all correct answers) [20].

A technology researcher (KH) was present throughout the study to manage the questionnaires and interviews as well as any technical issues that arose. This person did not intervene in the conduct of examination itself. A clinical researcher (MDS, JB, SMS, or CPK) was also available for patient or health professional queries.

Data Analysis

Videos were coded for task completion. A task was completed if all recordings were made, repeating and skipping movements as necessary, without intervention from the researcher. Support for system crashes was not counted as intervention because ASSESS MS is still at the prototype stage.

The length of examination was calculated for the first and final patients of each health professional. "Start time" was defined by the first keystroke of entering patient information and "end time" by the completion sound generated at the end of the final test, rounded to the nearest 5 seconds. Crashes were subtracted from the overall time from the moment the health professional realized there was a crash to the time at which he/she was able to resume. If the first or final patient was wheelchair bound or had severe cognitive decline identified by the health professional which changed the length of the examination substantially (because multiple movements were not performed), the next examination was used instead.

To test the statistical difference between length of examination of first and final patients of a health professional and between neurologists and nurses, we used Student *t* tests. Three videos could not be coded due to recording errors, such as missing beginning or video camera pointing in the wrong direction.

A metric of standardized movement performance was calculated from review of the depth videos. All sitting movements were scored for correct directionality of movement, for example, finger-nose test was performed with the arm to the side rather than front, and the correct number of repetitions (see Multimedia Appendix 3). Standing movements were not rated due to poor-quality images and/or preprocessing elements (eg, head detection) not yet available. Two people (CM and her colleague) rated 27 examinations, sampled to exclude the first 2 examinations by the participating health professionals, which were treated as training cases. Disagreements were resolved through discussion. Whether the camera view was clear was coded at the same time as the patient being visible throughout the entire examination.

The metric for adequate positioning was derived mathematically during the preprocessing of the depth videos before their usage in the machine-learning algorithms. In the case of lateral positioning, we report the number of pixels needed for the head to be transposed for it to be centered in the image. We considered anything within half the diameter of the "average" head size as "good positioning," as this would enable easy head detection. The "average" head size in our sample was 56 pixels. It was calculated by measuring the segmented area on 100 randomly picked processed finger-nose test videos at one fourth of the image height from the top. This should equate with about the eye level. For depth positioning, we report distance from that specified in the movement protocol. Any value within the diameter of the "average" head size (18 cm) was labeled as good positioning. This approach accounted for natural movement of the head.

The Likert-scale questionnaire data were tabulated and descriptive statistics were used. We considered answers per site as well as answers in aggregate; Student *t* tests were used to compare between sites. One patient's questionnaire data were removed, as this patient provided the same answer for all questions, suggesting a lack of attention to the questions. As the questions were balanced, both positive and negative answers would be expected. The responses to the free response questions (Q7-Q9) were minimal with only 46.4% (71/153) containing answers. The majority of answers contained only a few words that were prosaic in nature, for example,"Interesting system." These have been included in the reporting of the interviews when applicable but, for the most part, contributed little to the data analysis.

The words from the first free response question asking for 3 words to describe ASSESS MS to another patient were grouped into the following 3 categories: positive (eg, interesting), negative (eg, slow), and characteristic (eg, computer system). To gain a more nuance view of how the patients viewed ASSESS MS, a researcher (CM) and 2 visual designers grouped the words provided into a word cloud. These were melded into a visualization by an experienced visual designer. Words that are larger were those repeated more often (the number is shown in the visualization). Words on opposite sides of a line were considered contrasting. No words were deleted from the visualization so that the viewers might interpret for themselves the kind of language used to describe ASSESS MS.

The health professional interviews were coded for implicit or explicit discussions of standardization, patient-health professional interaction, and system comments. These themes were chosen after a first listening of the interviews as they encompassed the data, while providing responses to important design decisions. The interviews were further understood by viewing the associated examination videos after listening to the interviews.

Results

Participants

We recruited a convenience sample of 12 health professionals. Half were neurologists and half nurses. Participating health professionals were evenly split across 3 hospital sites in 2 countries. Their specialty and years of experience are presented in Table 2.

A total of 51 patients were recruited to the study. Slight over-recruitment (proposed sample n=48) was intended to address dropout due to patients choosing not to participate on the day; however, all patients recruited participated. Patients spanned a wide range of levels of physical and cognitive disability, as well as age, as seen in Table 3.

Table 2. Health professional characteristics.

Characteristics	Neurologists (n=6)	Nurses (n=6)
Age, mean (range), years	36.5 (26-53)	43.8 (27-61)
Gender (female/male)	4/2	6/0
Experience with MS, mean (range), years	4.3 (0.5-15)	2.7 (0.5-5)
Experience with physical examination, mean (range), years	3.4 (0.5-15)	0.8 (0-4)
Professional status	1 consultant, 2 attending physicians, 2 residents, and 1 medical student	4 study nurses, 1 clinical epidemiologist, and 1 study coordinator

Table 3. Study patient characteristics.

Characteristics	Total patients(n=51)
Age, mean (range), years	46.0 (23-73)
Gender (female/male)	31/20
Disease duration, mean (range), years	14.2 (0.5-47)
Disease course (CIS/RRMS/SPMS/PPMS)[a]	1/37/7/6
EDSS, median (range)	3 (1-7)
Symbol Digit Modalities Test, median (range)	47 (13-79)

[a]CIS = clinically isolated syndrome; PPMS=primary progressive MS; RRMS=relapsing remitting MS; SPMS=secondary progressive MS.

Usability

Effectiveness

All health professionals were able to carry out the examination appropriately using ASSESS MS without guidance after the first examination. This included positioning ASSESS MS, playing the instructional videos, capturing recordings, or navigating to different tests and subtests to repeat or skip a movement. No consistent task errors were identified by either researcher (CM/KH) involved in the analysis. Movements were performed by patients according to the protocol in 97.6% (405/415) of the cases. Most mistakes occurred in the drawing squares movement, but were not attributable to specific patients or health professionals. The camera view was never blocked. As illustrated in Table 4, no differences were seen across clinics in any metric. There was no substantive difference between neurologists and nurses.

Table 4. Metrics of usability (effectiveness) for each clinic and the aggregate.

Metric(patients)	Clinic 1(n=17)	Clinic 2(n=18)	Clinic 3(n=16)	Total(n=51)	Percentage
Task completion	17/17	18/18	16/16	51/51	100
Standardized movement performance	142/143	150/155	113/117	405/415	97.6
Clear camera view	17/17	18/18	16/16	51/51	100

Patients were laterally positioned consistently, with only 1 (n=368) outside the 28.5-pixel margin. There was less consistency in the distance between recording tool and patient, with 134 (n=368) being more than the 18 cm from the normalized distance. There was a substantial skew of videos being farther away than the requisite distance, as shown in Figure 5.

The questionnaire data indicate that both patients and health professionals gave high scores for ease of use as well as clarity of instruction as shown in Table 5. Patients gave higher ratings than health professionals. That said, many of the health professionals interviewed highlighted that the clear instructions given by the recording tool and minimal instruction given by the health professional should be effective in capturing data. As one health professional said:

“The system makes it easier to explain to the patient and they are more likely to do it correctly.” (HP7: Doctor)

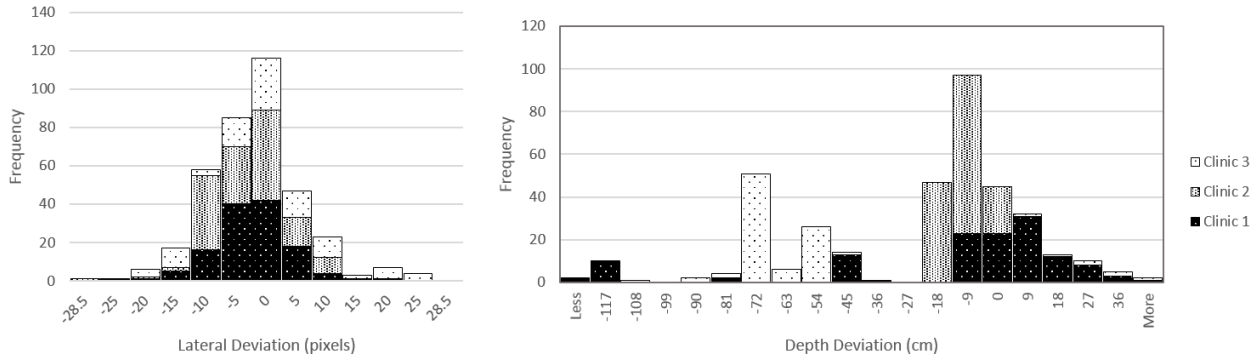
There were some differences between clinics, with Clinic 1 having significantly lower (P<.001) ratings than Clinic 2 by health professionals.

Table 5. Questionnaire data for effectiveness metric.

		Clinic 1	Clinic 2	Clinic 3	Total	Ideal[a]
Patients						
	I understood what to do during the study examination, mean (SD)	6.3 (1.3)	6.8 (0.4)	6.7 (0.5)	6.6 (0.9)	7
	The movement instructions given by the recording system were clear, mean (SD)	6.1 (1.4)	6.6 (0.6)	6.6 (0.5)	6.4 (1.0)	7
Health professionals						
	The recording system was easy to use, mean (SD)	5.0 (1.1)	6.7 (0.6)	5.4 (1.6)	5.7 (1.4)	7
	The movement instructions given by the recording system were clear to the patient, mean (SD)	5.1 (1.4)	6.2 (0.9)	5.7 (1.2)	5.7 (1.2)	7

[a]1 indicates “strongly disagree” and 7 indicates “strongly agree” (Likert scale).

Figure 5. Histogram of frequency of deviation from normalized point. Lateral deviation is presented in pixels (left) and depth deviation in centimeter (right).



Efficiency

The mean time to completion for a health professional’s final examination was 18:59 (mm:ss). There was a significant decrease in average recording time between the first and final examination times (P<.001), as shown in Table 6. This suggests that efficiency increases quickly with minimal experience. There were no significant differences (First examination: P=.55; Last examination: P=.91) in the length of examination between neurologists and nurses.

The recording tool was specifically highlighted as efficient by a number of health professionals. There was no comment on it being inefficient. As one health professional said,

“It is quick because it is so structured. What often happens is that the patient starts explaining things and talking about problems and it takes ways longer. If you follow it, you’re finished in 15-20 minutes. It keeps the focus on what matters.” (HP12: Nurse)

Table 6. Mean examination length of the first and final examinations of all health professionals and of nurses and doctors separately.

	First examinationMean (SD) time (mm:ss)	Final examinationMean (SD) (mm:ss)
Health professionals	26:55 (04:33)	18:59 (02:50)
Doctors	27:45 (04:28)	18:52 (03:50)
Nurses	26:06 (04:53)	19:05 (01:41)

Acceptability

Future Use

Patients disagreed with the statement that they would not like their health professional to use this system in future suggesting that it would be acceptable to use. Health professionals had a more varied view, but with aggregate scores that would still suggest acceptability (Table 7). One health professional opposed future use and gave very low scores. With no examination experience, health professionals found it challenging to manage both the patient and the technology, despite completing the recordings successfully.

Table 7. Questionnaire data for acceptability metric (future use).

		Clinic 1	Clinic 2	Clinic 3	Total	Ideal[a]
Patients						
	I would not like my health professional to use the recording system during my future examinations, mean (SD)	2.5 (1.7) (n=17)	2.0 (2.0) (n=18)	1.5 (1.5) (n=16)	2.0 (1.8) (n=51)	1
Health professionals						
	I would use the recording system in future examinations, mean (SD)	4.1 (1.4) (n=4)	6.8 (0.7) (n=4)	4.8 (1.2) (n=4)	5.2 (1.6) (n=12)	7

[a]1 indicates "strongly disagree" and 7 indicates "strongly agree" (Likert scale).

Human Interaction

The recording tool did not make either patients or health professionals feel awkward. The highest ambivalence came from patients and health professionals with regard to the instructional aspects of the recording tool. These 2 questions had the highest level of variability with a greater number of patients providing a neutral answer (Table 8).

Table 8. Questionnaire data for acceptability metric (interaction).

		Clinic 1	Clinic 2	Clinic 3	Total	Ideal[a]
Patients						
	I prefer my health professional to demonstrate the movements, mean (SD)	2.5 (1.8) (n=17)	1.6 (1.2) (n=18)	2.7 (1.7) (n=16)	2.3 (1.6) (n=51)	1
	The recording system made me feel awkward or uncomfortable, mean (SD)	1.8 (1.3) (n=17)	1.8 (1.7) (n=18)	1.1 (0.3) (n=16)	1.6 (1.4) (n=51)	1
Health professionals						
	I prefer to demonstrate the movements to the patient myself, mean (SD)	3.7 (1.7) (n=4)	2.3 (1.6) (n=4)	4.1 (1.0) (n=4)	3.4 (1.7) (n=12)	1
	The recording system made me feel awkward or uncomfortable, mean (SD)	2.1 (0.2) (n=4)	1.4 (0.8) (n=4)	2.4 (1.4) (n=4)	2.0 (1.0) (n=12)	1

[a]1 indicates "strongly disagree" and 7 indicates "strongly agree" (Likert scale).

Early testing of the prototype with neurologists suggested that some felt that ASSESS MS usurped their role, as articulated in the following quotation, and might pose an issue to acceptability:

"Usually everything that the computer tells the patient is something that you tell the patient as a physician, so the interaction is somehow reduced because you let the computer talk...I am used to tell the patient exactly what I want them to do so I can see what I want to see. There is nothing a physician needs to do. It is something that my assistant could do." (Doctor)

We found similar sentiments in this study as well, formulated in different ways. One neurologist reflected on the feeling of loss of the physical connection with the patient that would normally be gained through touching the patient during the examination. Another neurologist discussed the disruption to her rhythm, saying "I have my rhythm, an interaction with just me and the patient. Here we have the third component...It's a threesome." A third neurologist spoke about an inability to move freely about the room as ASSESS MS occupied the limited floor space. A fourth neurologist spoke about the loss of the creative process of medicine, suggesting that this test could be done by assistants.

Despite these initial discomforts mentioned, neurologists adapted quickly to engaging with the patient while using ASSESS MS. Several neurologists noted that they said more than they needed to in the beginning, but decreased their verbal speech with time. As one said,

"The last patients understood the instructions better so I did not say anything and they did it well." (HP11: Doctor)

Others felt that an examination was more personal if they spoke more often.

"I think I explain a bit more than necessary...Sometimes I just repeat what has already been said [by ASSESS MS]. I think sometimes the patient would have been able to just understand it just by listening. I personally think that it is somehow more personal if I say it again or point out what could have been important." (HP6: Doctor)

In all cases, the doctors spoke less over time, with many using substantial body language, such as exaggerated nods or smiles to replace verbal interaction. Reflection on the videos suggests that all doctors had found an examination rhythm by the third patient. This was often gained through pre-emptively instructing the patients in the aspects of the movements that needed to be standardized.

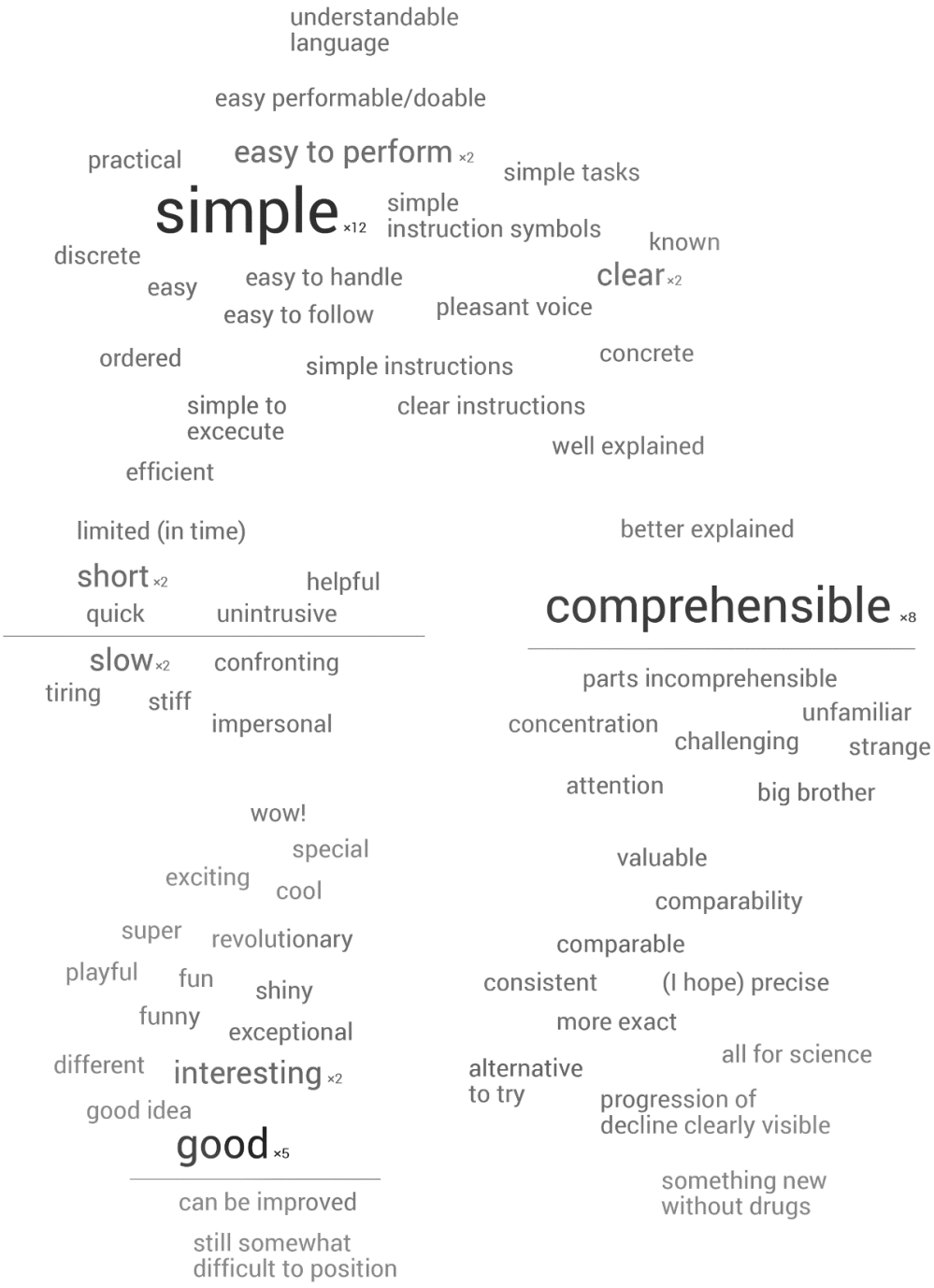
Nurses did not have the same feelings and raised no comparable comments about using ASSESS MS. They took a different

perspective on the instructional videos. Most mentioned the dual role of the video in advising them as well as the patient as to what to do. This is illustrated by requests from 3 nurse participants to have protocol information verbally included in the movement descriptions, such as “With the feet on the floor, raise your arm out to the side...” The nurses easily found a balance between interaction with the patient and ASSESS MS.

Perceptions

Patient and health professionals were overwhelmingly positive when asked to give 3 words describing ASSESS MS. Upon tabulating the individual words patients used to describe ASSESS MS, 78 were found to be positive (eg, simple), 12 negative (eg, slow), and 29 characteristic (eg, computer system). For the health professionals, there were 19 positive, 2 negative, and 5 characteristic words. Figure 6 provides a visualization of all of the words used by patients to describe ASSESS MS. The feedback of the health professionals focused mainly on specific technical fixes (eg, the improved phrasing of a particular movement instruction). No negative feedback was given during the interviews.

Figure 6. Word cloud of patientsâ€™ descriptive words of ASSESS MS.



Discussion

Principal Findings

The ASSESS MS is a system to support the assessment of motor dysfunction in patients with MS using depth-sensing computer vision. It aims to provide a consistent, quantified metric of motor ability to enable finer-grained tracking of disease progression than currently possible. ASSESS MS has been designed to facilitate more standardized data capture to support movement analysis by the machine-learning algorithms, while being both usable and acceptable by patients and health professionals in clinic settings. Our results show high levels of usability and acceptability by both patients and health professionals. There was little variation in results across the 3 participating centers, and no differences between neurologists and nurses.

The study suggests that ASSESS MS is usable. It is effective in that all health professionals were able to complete the recordings, with high levels of standardization of movement performance, lateral positioning, and clear camera view. It was also perceived as easy to use by both patients and health professionals with high scores on the Likert-scale questions and positive comments being provided during the interviews. The variation achieved in the patient population indicates that patients' level of physical (EDSS range from 1 to 7) or cognitive (SDMT from 13 to 79) disability does not change the effectiveness of the tool.

The only aspect of ASSESS MS that was problematic was the distance positioning. Closer inspection of the videos suggested that distance was frequently gauged through physical landmarks in the room (eg, wall), rather than the distance provided on screen. Clinic 2, which had room furniture at the correct distances to facilitate alignment had the highest consistency of depth positioning, placing ASSESS MS in front of a bookcase for the sitting exercises and then against a desk for the standing exercises. This suggests that the physical properties of the room should be reviewed and highlighted in the training when a new site is trained to use ASSESS MS.

ASSESS MS also seemed to be reasonably efficient with the average time under 20 minutes and decreasing with use. Because the examination did not incorporate the complete EDSS, no direct comparison was possible. The lower bound of completion time, about 13 minutes, suggests that the test may be completed in less time with further emphasis on speed. That health professionals felt it was efficient is also important.

ASSESS MS was acceptable to patients on all dimensions considered, including future use, interaction (with health professionals), and perceptions. There was little variation of scores, with most clustered at one end of the Likert-scale and with a very few at the opposite end. This suggests that most people had high acceptability and a small number would not use it, but there was little ambivalence. There was no clear pattern of those who opposed it, for example, by age or level of disability. Variations in attitudes to new technologies, including negative perceptions, are predicted by models of staged technology adoption [21]. The word descriptions of ASSESS MS were overwhelmingly positive. These results strongly suggest that ASSESS MS is acceptable to patients.

Health professionals also accepted ASSESS MS, but with greater ambivalence. Clinic 1 seemed to have particularly low scores, which we attribute to experiencing the most technical issues, such as the disconnection of the Kinect camera feed from the application that required a restart by the supervising technology researcher (KH). It was interesting that the 2 professionals who were new to MS had no difficulty performing the test with movements that they did not know. Two other health professionals with little or no examination experience in any area of medicine, however, were more uncomfortable. Training in examination skills should also be provided for those without previous experience.

There were no differences between neurologists and nurses in any quantitative metric analyzed, including length of examination. The only difference lay in attitudes. Neurologists found it initially uncomfortable to work with this semiautomated, highly standardized system, whereas nurses welcomed the support the system provided. That said, neurologists used their examination skills to pre-empt potentially incorrectly performed movements building an interaction around achieving standardized movement performance. We would suggest that ASSESS MS is particularly well suited to health professionals other than neurologists, but is flexible enough to be used by any health professional.

These findings indicate that ASSESS MS in its current version is both usable and acceptable. It can be deployed to new sites and used by a range of health professionals with just 1 hour of training. This stands in contrast to current tools, such as the EDSS, which require a standardized training and an experienced background in clinical neurology. As such, ASSESS MS has the potential for inexpensive, widespread use.

Anecdotal Lessons

The training process gave substantial insight into how health professionals came to understand ASSESS MS. It worked best to provide a simple characterization of how the machine learning worked as a process of comparison between new data and past patients that it had "seen." This meant that if a patient "looked like" they had a given disability level, then they would be labeled as such, irrespective of whether that was because they actually had that level of dysfunction or because they did not perform a movement correctly. Providing this fairly simple account of how ASSESS MS works was motivating for the health professionals in trying to achieve standardized data.

We also found that a small change to the way the movement protocol was introduced, by asking health professionals to perform it as opposed to watch it, increased confidence. That said, the movement protocol was the most challenging part of the health professionals' learning curve, and perhaps not captured in the system-oriented measures in this paper. There were numerous questions about how to perform the protocol and the availability of one of the clinical researchers to correct errors between examinations was welcomed. When considering training in future, the availability of such a person should be

included, and the greater emphasis on the movements as opposed to system use should be considered.

Limitations

Usability is only generally specified, with specific metrics required for each technology being assessed. We could have assessed many aspects of ASSESS MS, but have focused on ones that we think are critical to producing a workable system for both health professionals and data engineers. As a novel technology that aims to achieve something very different than the status quo, a direct comparative is not available. As such, what constitutes a "good" result can be disputed. It is difficult, for example, to provide an exact bound on what is acceptable positioning or a necessary percentage of standardized movement. That said, we know that these elements are essential to the system and want to continue increasing these numbers.

The choices made for measures matched the design criteria of ASSESS MS, but technology moves quickly. The original Kinect is no longer on the market, for example, and the machine-learning algorithms will continue to develop. Further, ASSESS MS focuses only on a subpart of current disability measurements and some of our results (eg, standardized movement performance) apply only to sitting movements. While the specific results offered in this paper give insight into usage and perception that are unlikely to change dramatically, the quality of the data produced needs to be continually assessed. New techniques to continually evaluate a system as changes are introduced are needed to provide sustained evidence of usability [22].

In addition, no testing was performed with participants with severe visual or hearing loss, although we would expect the health professional to play a mediating role in these situations if extra help is required.

Comparison With Previous Work

There are a growing number of computer-assisted and sensor-based applications that support clinical assessment and rehabilitation for MS patients. These include social gaming [23], exoskeletons [24], and virtual environments [25]. More recently, a number of depth-sensing computer vision applications for MS rehabilitation have also been evaluated (eg, [7]), which show both successful implementation and good acceptability by patients.

Sensor-based recordings of human movements are becoming increasingly important for assessment of symptoms in different neurological disorders in addition to the strides made in rehabilitation [26]. In MS specifically, body-worn motion sensors can detect mobility difference between healthy volunteers and patients with early stage MS better than traditional time tests [27]. Accelerometers have also been used to measure both physical activity and walking mobility in MS patients [28]. Most recently, an accelerometer built into an iPad has been used for gait and balance analysis [29]. There are also initial findings about the use of depth sensing for carrying out gait analysis in MS patients [30]. Sensing technology to support patients with other conditions that cause motor dysfunction is also being developed [31].

Despite initial research in the area, applications that use sensing (vision or other) for clinical assessment have not been deployed in clinical settings. This study shows that attention to the clinical environment in the design process can make these new approaches to medicine a reality.

Conclusions

Depth-sensing computer vision has been rapidly adopted to form the core of a range of innovative health care applications for the clinical assessment and rehabilitation of movement ability. There are now increasing numbers of examples of the commercialization of such ideas for rehabilitation. Yet, clinical assessment has been a greater challenge, with the need for greater precision of measure, and hence lower variability in the data.The creation of ASSESS MS, as part of one of the first projects in this domain, shows that careful attention to deployment makes it possible to collect sensor data of a quality needed for clinical assessment. Moreover, it can be done in a way that is suitable to wide-scale deployment and acceptable to patients. These results open the door for greater development in depth-sensor-based assessment of movement disorders.

Acknowledgments

We would like to acknowledge the participation of the patients and health professionals who took part in the study. Specifically, Karin Streit, Monika Käser, Andrea Seiler, Verena Blatter, Cyra Leurs, Lisanne Balk, Tim Alleman, Danielle Savelkoel, Edwige Bolliger, Suzy Gajewski, Gerlinde Hoffmann, Jakob Stockinger. We would like to thank Michela Azzarito for providing the video coding tool. The research has been funded by Novartis Pharma AG, CH-4002 Basel, Switzerland, who have been active partners in the technology development process. There has been no restriction placed on the researchers involved.

Authors' Contributions

CM, KH, and AS designed the study with clinical input from MDS, CPK, SMS, JB, and FD and technical input from PK, JFD, and AC. The paper was drafted by CM, KH, and MDS. BU and LK took part in the general leadership of this project. All authors took part in the revision of this paper.

Conflicts of Interest

CM, KH, PK, AC, and AS are/were employees of Microsoft Research. FD and JFD are employees of Novartis Pharma AG. MDS, CPK, SMS, JB, BU, and LK have no conflicts of interest.

Multimedia Appendix 1

ASSESS MS training materials.

[PDF File (Adobe PDF File), 411KB - humanfactors_v2i1e11_app1.pdf]

Multimedia Appendix 2

Questionnaire and interview materials.

[PDF File (Adobe PDF File), 201KB - humanfactors_v2i1e11_app2.pdf]

Multimedia Appendix 3

Standardized movement performance rating guide.

[PDF File (Adobe PDF File), 90KB - humanfactors_v2i1e11_app3.pdf]

References

1. Simon SR. Quantification of human motion: Gait analysis-benefits and limitations to its application to clinical problems. J Biomech 2004 Dec;37(12):1869-1880. [doi: 10.1016/j.jbiomech.2004.02.047] [Medline: 15519595]
2. O'Hara K, Gonzalez G, Penney G, Sellen A, Corish R, Mentis H, et al. Interactional order and constructed ways of seeing with touchless imaging systems in surgery. Comput Support Coop Work 2014 May 7;23(3):299-337. [doi: 10.1007/s10606-014-9203-4]
3. Omelina L, Jansen B, Bonnechère B, S Van Sint Jan J, Cornelis J. Serious games for physical rehabilitation: Designing highly configurable and adaptable games. 2012 Presented at: Proceedings of the 9th International Conference on Disability, Virtual Reality and Associated Technologies (ICDVRAT); Sept 10-12, 2012; Laval, France.
4. Kayama H, Okamoto K, Nishiguchi S, Yamada M, Kuroda T, Aoyama T. Effect of a Kinect-based exercise game on improving executive cognitive performance in community-dwelling elderly: Case control study. J Med Internet Res 2014 Jan;16(2):e61 [FREE Full text] [doi: 10.2196/jmir.3108] [Medline: 24565806]
5. Eguíluz G, García MB. Use of a time-of-flight camera with an Omek Beckon framework to analyze, evaluate and correct in real time the verticality of multiple sclerosis patients during exercise. Int J Environ Res Public Health 2013 Nov;10(11):5807-5829 [FREE Full text] [doi: 10.3390/ijerph10115807] [Medline: 24192790]
6. Ortiz-Gutiérrez R, Cano-de-la-Cuerda R, Galán-del-Río F, Alguacil-Diego IM, Palacios-Ceña D, Miangolarra-Page JC. A telerehabilitation program improves postural control in multiple sclerosis patients: A Spanish preliminary study. Int J Environ Res Public Health 2013 Nov;10(11):5697-5710 [FREE Full text] [doi: 10.3390/ijerph10115697] [Medline: 24185843]
7. Lozano-Quilis J, Gil-Gómez H, Gil-Gómez J, Albiol-Pérez S, Palacios-Navarro G, Fardoun HM, et al. Virtual rehabilitation for multiple sclerosis using a Kinect-based system: Randomized controlled trial. JMIR Serious Games 2014;2(2):e12 [FREE Full text] [doi: 10.2196/games.2933] [Medline: 25654242]
8. Morrison C, Culmer P, Mentis H, Pincus T. Vision-based body tracking: Turning Kinect into a clinical tool. Disabil Rehabil Assist Technol 2014 Dec 11:1-5. [doi: 10.3109/17483107.2014.989419] [Medline: 25495708]
9. Bonnechère B, Jansen B, Salvia P, Bouzahouene H, Omelina L, Moiseev F, et al. Validity and reliability of the Kinect within functional assessment activities: Comparison with standard stereophotogrammetry. Gait Posture 2014 Jan;39(1):593-598. [doi: 10.1016/j.gaitpost.2013.09.018] [Medline: 24269523]
10. Mentiplay BF, Clark RA, Mullins A, Bryant AL, Bartold S, Paterson K. Reliability and validity of the Microsoft Kinect for evaluating static foot posture. J Foot Ankle Res 2013 Jan;6(1):14 [FREE Full text] [doi: 10.1186/1757-1146-6-14] [Medline: 23566934]
11. Stone EE, Skubic M. Unobtrusive, continuous, in-home gait measurement using the Microsoft Kinect. IEEE Trans Biomed Eng 2013 Oct;60(10):2925-2932. [doi: 10.1109/TBME.2013.2266341] [Medline: 23744661]
12. Caspar P, Otte K, Mansow-Model S, Paul F, Brandt A. Kinect-based analysis of posture, gait and coordination in multiple sclerosis patients. Neurology 2013;80(1001):P04.097.
13. Kamm CP, Uitdehaag BM, Polman CH. Multiple sclerosis: Current knowledge and future outlook. Eur Neurol 2014;72(3-4):132-141 [FREE Full text] [doi: 10.1159/000360528] [Medline: 25095894]
14. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). Neurology 1983 Nov 01;33(11):1444-1452. [doi: 10.1212/WNL.33.11.1444] [Medline: 6685237]
15. Cohen JA, Reingold SC, Polman CH, Wolinsky JS. Disability outcome measures in multiple sclerosis clinical trials: Current status and future prospects. Lancet Neurol 2012 May;11(5):467-476. [doi: 10.1016/S1474-4422(12)70059-5] [Medline: 22516081]
16. Kontschieder P, Dorn JF, Morrison C, Corish R, Zikic D, Sellen A, et al. Quantifying progression of multiple sclerosis via classification of depth videos. Med Image Comput Comput Assist Interv 2014;17(Pt 2):429-437. [Medline: 25485408]

17. Chua P, Crivella R, Daly B, Hu N, Schaaf R, Ventura D, et al. Training for physical tasks in virtual environments: Tai Chi. New York: IEEE; 2003 Mar 22 Presented at: IEEE Virtual Reality Conference 2003; Mar 22-26, 2003; Los Angelos, CA p. 87-94. [doi: 10.1109/VR.2003.1191125]
18. Pomplun M, MatariÄ‡ M. Evaluation metrics and results of human arm movement imitation. 2000 Sep 07 Presented at: International Conference on Humanoid Robots (Humanoids 2000); Sept 7-8, 2000; Cambridge, MA.
19. International Organization for Standardization. Ergonomic Requirements for Office Work With Visual Display Terminals (VDTs). Part 11: Guidance on Usability (ISO 9241-11:1998). Geneva, Switzerland: International Organization for Standardization; 1998. URL: http://www.iso.org/iso/catalogue_detail.htm?csnumber=16883 [accessed 2015-06-15] [WebCite Cache ID 6ZJ23kOY6]
20. Smith A. Symbol Digit Modalities Test: Manual. San Francisco, CA: Western Psychological Corporation; 2002.
21. Holden RJ, Karsh BT. The technology acceptance model: Its past and its future in health care. J Biomed Inform 2010 Feb;43(1):159-172. [doi: 10.1016/j.jbi.2009.07.002] [Medline: 19615467]
22. Morrison C, Doherty G. Analyzing engagement in a web-based intervention platform through visualizing log-data. J Med Internet Res 2014;16(11):e252 [FREE Full text] [doi: 10.2196/jmir.3575] [Medline: 25406097]
23. Weyer TD, Robert K, Hariandja J, Alders G, Conin K. The social maze: A collaborative game to motivate MS patients for upper limb training. Berlin, Germany: Springer; 2012 Sep 26 Presented at: International Conference on Entertainment Computing (ICEC 2012); Sept 26-28, 2012; Bremen, Germany p. 476-479. [doi: 10.1007/978-3-642-33542-6_57]
24. Gijbels D, Lamers I, Kerkhofs L, Alders G, Knippenberg E, Feys P. The Armeo Spring as training tool to improve upper limb functionality in multiple sclerosis: A pilot study. J Neuroeng Rehabil 2011 Jan 24;8:5 [FREE Full text] [doi: 10.1186/1743-0003-8-5] [Medline: 21261965]
25. Baram Y, Miller A. Virtual reality cues for improvement of gait in patients with multiple sclerosis. Neurology 2006 Jan 24;66(2):178-181. [doi: 10.1212/01.wnl.0000194255.82542.6b] [Medline: 16434649]
26. Khusainov R, Azzi D, Achumba IE, Bersch SD. Real-time human ambulation, activity, and physiological monitoring: Taxonomy of issues, techniques, applications, challenges and limitations. Sensors (Basel) 2013 Sep;13(10):12852-12902 [FREE Full text] [doi: 10.3390/s131012852] [Medline: 24072027]
27. Spain RI, St George RJ, Salarian A, Mancini M, Wagner JM, Horak FB, et al. Body-worn motion sensors detect balance and gait deficits in people with multiple sclerosis who have normal walking speed. Gait Posture 2012 Apr;35(4):573-578 [FREE Full text] [doi: 10.1016/j.gaitpost.2011.11.026] [Medline: 22277368]
28. Weikert M, Motl RW, Suh Y, McAuley E, Wynn D. Accelerometry in persons with multiple sclerosis: Measurement of physical activity or walking mobility? J Neurol Sci 2010 Mar 15;290(1-2):6-11. [doi: 10.1016/j.jns.2009.12.021] [Medline: 20060544]
29. Rudick RA, Miller D, Bethoux F, Rao SM, Lee J, Stough D, et al. The Multiple Sclerosis Performance Test (MSPT): An iPad-based disability assessment tool. J Vis Exp 2014 Jun 30(88):e51318 [FREE Full text] [doi: 10.3791/51318] [Medline: 25046650]
30. Behrens J, Otte K, Mansow-Model S, Brandt A, Paul F. Kinect-based gait analysis in patients with multiple sclerosis. Neurology 2014 Apr 29;82(10):P3.135.
31. TakaÄ B, Català A, Rodríguez Martín D, van der Aa N, Chen W, Rauterberg M. Position and orientation tracking in a ubiquitous monitoring system for Parkinson disease patients with freezing of gait symptom. JMIR Mhealth Uhealth 2013;1(2):e14 [FREE Full text] [doi: 10.2196/mhealth.2539] [Medline: 25098265]

Abbreviations

EDSS: Expanded Disability Status Scale
MS: multiple sclerosis
SDMT: Symbol Digit Modalities Test

Edited by G Eysenbach; submitted 23.01.15; peer-reviewed by B Jansen, M Chary; comments to author 06.04.15; revised version received 09.04.15; accepted 07.05.15; published 24.06.15.

Please cite as:
Morrison C, D'Souza M, Huckvale K, Dorn JF, Burggraaff J, Kamm CP, Steinheimer SM, Kontschieder P, Criminisi A, Uitdehaag B, Dahlke F, Kappos L, Sellen A
Usability and Acceptability of ASSESS MS: Assessment of Motor Dysfunction in Multiple Sclerosis Using Depth-Sensing Computer Vision
JMIR Human Factors 2015;2(1):e11
URL: http://humanfactors.jmir.org/2015/1/e11/
doi:10.2196/humanfactors.4129
PMID:27025782

Original Paper

Enhancing the Effectiveness of Consumer-Focused Health Information Technology Systems Through eHealth Literacy: A Framework for Understanding Users' Needs

Lars Kayser[1*], MD, PhD; Andre Kushniruk[2*], PhD; Richard H Osborne[3*], PhD; Ole Norgaard[1*], MSc; Paul Turner[4*], PhD

[1]Department of Public Health, University of Copenhagen, Copenhagen, Denmark

[2]School of Health Information Science, University of Victoria, Victoria, BC, Canada

[3]Deakin Population Health Strategic Research Centre, Deakin University, Melbourne, Australia

[4]eHealth Services Research Group, University of Tasmania, Hobart, Australia

[*]all authors contributed equally

Corresponding Author:
Lars Kayser, MD, PhD
Department of Public Health, University of Copenhagen
Øster Farimagsgade 5
Copenhagen, 1014
Denmark
Phone: 45 28757291
Fax: 45 35327285
Email: lk@sund.ku.dk

Abstract

Background: eHealth systems and applications are increasingly focused on supporting consumers to directly engage with and use health care services. Involving end users in the design of these systems is critical to ensure a generation of usable and effective eHealth products and systems. Often the end users engaged for these participatory design processes are not actual representatives of the general population, and developers may have limited understanding about how well they might represent the full range of intended users of the eHealth products. As a consequence, resulting information technology (IT) designs may not accommodate the needs, skills, cognitive capacities, and/or contexts of use of the intended broader population of health consumers. This may result in challenges for consumers who use the health IT systems, and could lead to limitations in adoption if the diversity of user attributes has not been adequately considered by health IT designers.

Objective: The objective of this paper is to propose how users' needs and competences can be taken into account when designing new information and communications technology solutions in health care by expanding the user-task-context matrix model with the domains of a new concept of eHealth literacy.

Methods: This approach expands an existing method for supporting health IT system development, which advocates use of a three-dimensional user-task-context matrix to comprehensively identify the users of health IT systems, and what their needs and requirements are under differing contexts of use. The extension of this model involved including knowledge about users' competences within the seven domains of eHealth literacy, which had been identified based on systematic engagement with computer scientists, academics, health professionals, and patients recruited from various patient organizations and primary care. A concept map was constructed based on a structured brainstorm procedure, card sorting, and computational analysis.

Results: The new eHealth literacy concept (based on 7 domains) was incorporated as a key factor in expanding the user-task-context matrix to describe and qualify user requirements and understanding related to eHealth literacy. This resulted in an expanded framework and a five-step process, which can support health IT designers in understanding and more accurately addressing end-users' needs, capabilities, and contexts to improve effectiveness and broader applicability of consumer-focused health IT systems. It is anticipated that the framework will also be useful for policy makers involved in the planning, procuring, and funding of eHealth infrastructure, applications, and services.

Conclusions: Developing effective eHealth products requires complete understanding of the end-users' needs from multiple perspectives. In this paper, we have proposed and detailed a framework for modeling users' needs for designing eHealth systems that merges prior work in development of a user-task-context matrix with the emerging area of eHealth literacy. This framework

is intended to be used to guide design of eHealth technologies and to make requirements explicitly related to eHealth literacy, enabling a generation of well-targeted, fit-for-purpose, equitable, and effective products and systems.

(JMIR Human Factors 2015;2(1):e9) doi:10.2196/humanfactors.3696

KEYWORDS

eHealth literacy; requirements; user involvement

Introduction

A major factor identified as a point of failure in the development and implementation of health information systems is limited understanding of users, their needs, and the contexts in which the systems are used. Measurable benefits from involving users in the design and development of information technology (IT) have been demonstrated in some studies [1]. However, the comparison of outcomes across these studies is often problematic due to the different ways users are defined, engaged, and the degree to which they are involved in real-world settings. Importantly, this has tended to obscure the way in which different users' needs, capabilities, and contexts have an impact on the safety and usability of eHealth systems. Because eHealth systems increasingly focus on enabling autonomous use by health consumers, the implications of these differences for patient safety require even closer attention and investigation.

For health consumers, access to networked information technologies has primarily emerged as a way of either complementing or by-passing conventional sources of health-related information. These technologies are increasingly being used to share health information, personal experiences, and knowledge of medications and medical services, as well as to offer support and the tracking of personal health care [2]. The capacity of health consumers to successfully use these systems links back to the classical problem of communication and challenges arising from any misalignment in the communication triangle (the sender, the message, and the recipient) [3]. When health IT is used for information and/or communication (the message), outcomes rely heavily on an alignment of the system's (the sender's) capability to address and acknowledge the health consumer's (the recipient's) attributes. From the user perspective, the consumer needs to possess both the confidence and skills to acquire information, understand it, and actively appraise it [4]. When using digital media, this means that the consumer needs to be *eHealth literate* [5].

From a health IT design perspective, to ensure that systems have a high level of usability (ie, high efficiency, effectiveness, and generate satisfaction for users) [6], it is necessary that the needs, capabilities, and contexts of use of these users are known and incorporated into the IT design. For more than a decade, different approaches have been taken to illuminate the users' interaction with IT systems (eg, by assessing their digital literacy or confidence in these systems) [7-9]. In 2006, Norman and Skinner [10] proposed a model of eHealth literacy, taking six core literacies (ie, traditional, media, computer, information, science, and health literacy) into one concept, defined as "an individual's ability to search for, successfully access, comprehend, and appraise desired health information from electronic sources and to then use such information to attempt to address a particular health problem". They also developed the eHealth Literacy Scale (eHEALS) instrument [10] that can be used to measure eHealth literacy. However, transformation of health IT from simple and static websites (Web 1.0) to more complex and dynamic applications, online services, and social media (Web 2.0) has superseded the original conceptual base of the eHEALS and other instruments [11].

As private sector providers have entered the social media market to deliver support to patients online [12], other trends highlight the negative impacts of eHealth arising from misinformation [13] and/or reinforcement of negative behaviors [14], ultimately compromising the consumers' safety.

This calls for new ways for providers to address their consumers being aware of their needs and capabilities. In response to these challenges, we have developed a new concept of eHealth literacy to be reported in detail in due course.

This paper presents a framework grounded in a multidisciplinary approach that supports the modeling of user requirements and understanding of their needs by integrating a new concept of eHealth literacy with an expanded user-context-task matrix advocated for safe eHealth systems design.

The framework presented is structured as a five-step process aimed at supporting eHealth system designers to improve their understanding and more accurately address end-users' needs, capabilities, and contexts of use to improve the effectiveness and broader applicability of consumer-focused eHealth systems and applications. Using this framework during design will make requirements related to eHealth literacy explicit and contribute to the delivery of better-targeted, fit-for-purpose, equitable, and safe eHealth applications and systems to maximize consumer empowerment and health outcomes while reducing health inequalities. It is also anticipated that the framework will raise awareness of eHealth literacy issues among policymakers involved in the planning, procuring, and funding of eHealth infrastructure, applications, and services.

Methods

eHealth Literacy Concept Development

Although the focus of this paper is on presenting a framework for understanding users' needs to enhance eHealth systems design, it is useful to briefly describe the fieldwork approach that led to the development of the new concept of eHealth literacy used within this paper.

The key points of the new "eHealth literacy" concept are based on a highly structured and rigorous approach to questionnaire development—the validity-driven approach [15]. This is the same methodology that was used in the development of the

Health Literacy Questionnaire [16] and in numerous other widely used questionnaires [17-21].

Between June and August 2012, eight consultation sessions were conducted by two of the authors of this paper (ON and RHO) with participation of computer scientists, academics, health professionals, and patients recruited from patient organizations and primary care. During each session, a concept map was constructed based on a structured brainstorm procedure, statement generation, card sorting, and computational analysis [22]. A total of 458 statements were grouped into 68 clusters initially using hierarchal cluster analysis based on the participant sorting data undertaken in each session. Finally, qualitative synthesis was applied to reduce the clusters to smallest number of conceptually distinct concepts of eHealth literacy (Textbox 1). The results were confirmed by an Internet-based consultation with stakeholders.

The resulting seven domains were further grouped and reflected three overarching themes related to end users and technology. The first theme (capabilities) was the end-users' capabilities within the areas of health, information, and technology; the second theme (access to technologies) was the end-users' relationship to, or perceptions relating to, interacting with technology—more specifically, whether technology was perceived as being accessible and suits individual needs; and the third theme (experience using technologies) was the experience of benefits of using technology including being in control.

The concept is not only a template for an ongoing development of a new eHealth literacy instrument but also a model grounded in the experiences of modern-day IT users.

As a result, the seven domains of this eHealth literacy concept could be integrated with conventional IT design processes to enhance the understanding of users' needs among designers of eHealth systems and applications. The next section of the paper focuses on the integration of the new eHealth literacy concept (encapsulated in the seven domains described in Textbox 1) with the expanded user-task-context matrix.

Textbox 1. A new concept of eHealth literacy.

1. Capabilities

   Knowledge about one's own health (Domain 1)

   Know about the body's basic functions and structure and own current health status. Aware of risk factors and how to avoid them or reduce their influence on own health.

   Ability to interact with information (Domain 2)

   Able to read, write and remember, apply basic numerical concepts, and understand context-specific language (e.g., health, IT or the user's native language, as well as critically appraise information. Know when, how and what information to use.

   Ability to engage with technology (Domain 3)

   Being comfortable using computers and other digital media for handling information.

2. Access to technologies

   Access to technologies that work (Domain 4)

   Have access to technologies (e.g. computers and other digital media) that the users trust to be working *when* they need it and *as* they expect it to work.

   Access to technologies that suit individual needs (Domain 5)

   Have access to technologies that are adaptable to the specific needs and preferences of the users. This includes responsive features of both technologies and the healthcare system (including carers) as well as adaptation of devices and interfaces to be used by people with physical and mental disabilities.

3. Experience using technologies

   Feel that using technologies is beneficial (Domain 6)

   Feel that engaging in the use of technologies will help them to manage their health more effectively than by other means.

   Feel in control and secure when using technologies (Domain 7)

   Feel that you have the ownership of personal data stored in the systems and that the data are safe and can be accessed only by people to whom they are relevant (own doctor, own nurse etc.).

A Framework for Design: Extending the User-Task-Context Matrix With eHealth Literacy

Kushniruk and Turner [23] proposed a concept of a user-task-context matrix to help guide health IT system designers and developers regarding the gathering of user requirements, selection of end users for participatory design as well as for testing and evaluating resultant health IT systems (Figure 1).

Building on the conventional user-task matrix, a standard framework used in development of IT systems for modeling users and the tasks they are expected to be able to carry out using a new technology was developed previously [24]. Kushniruk and Turner [23] proposed that to adequately address

safety and quality concerns in health IT, a third dimension was required: the context of use of a health IT system. The original two-dimensional user-task matrix has been successfully applied in IT design and has been described in detail by Hackos and Redish [24] and is part of the human-computer interaction literature. However, in health care settings, in addition to consideration of the user and tasks, the context of use has a significant impact on whether a particular class of user can carry out specific tasks successfully or not using a particular device or information system. "Context" refers to the setting or conditions under which the technology is used; for example, is a system used under urgent or nonurgent conditions? Or is it used in a clinical setting, at home, at work, or while traveling? The "class of user" refers to the generic type of users of an eHealth application, that is, users who share basic characteristics (eg, users of a particular age group or patient users having a specific disease) [23,24]. While designing eHealth applications it is important to have adequate knowledge about all the major types (or classes) of users to target and tailor the eHealth application to meet the needs of each class of users (as different classes of users will have different information needs, understanding of health, and information-processing capabilities). The work of Kushniruk and Turner [23] extends and expands the concept of the user-task matrix to consider the key aspect of context of use. The user-task-context framework has also proven useful in the modeling of user requirements for different health care applications, mostly involving development of systems for health professionals [25,26].

In the context of the development of IT and applications for patients and other health consumers, it is especially important to add the concept of eHealth literacy to the user-task-context matrix in order to create an expanded framework. However, the existing concept of eHealth literacy proposed by Norman and Skinner [5] and subsequently elaborated by Chan and Kaufman [27] does *not* cover the consumer competences required for end users to benefit from and/or have the ability to interact meaningfully with contemporary eHealth systems.

However, with our new modern concept of eHealth literacy, it is now possible to integrate eHealth literacy and the user's competence into the user-task-context matrix, which makes it possible to generate a framework for health IT design that captures a diversity of consumer-related aspects.

The original empirical work that led to this new redefined eHealth literacy concept was intended to highlight to a broad group of potential stakeholders (policymakers, health care providers, clinicians, system designers) the diversity of end-user capability, access, and experience with IT systems. However, further analysis of the concept with regard to the user-task-context matrix revealed that it is possible to extend this matrix to produce a design framework that would enhance consideration of end users in consumer health IT design processes (Figure 2).

Figure 2 illustrates how the seven domains of the eHealth literacy concept relate to the user-task-context matrix model to generate a framework for design. User interactions with digital services always take place in a context of use, illustrated by the large outer circle. The use of digital services is mediated through tasks to be performed and these must be encoded by developers into health IT systems. The services of these health IT systems must, in turn, be intelligible to end users who require the knowledge and skills to find, understand, interpret, and decide how to respond to these outputs (domains 1, 2, and 3). More specifically, the framework highlights that domains 6 and 7 (in the intersection between users and tasks) rely on a detailed understanding of the factors that designers can use to motivate end users and enforce their behavior in the direction of being more self-directed, and domains 4 and 5 require designers to focus primarily on systems and tasks to ensure they can accommodate and support a variety of types of user access and diverse users' needs. Based on this expanded framework, it is proposed that the concept of eHealth literacy adds value to health IT design processes in at least three ways:

- During requirements gathering, an approach underpinned by principles can be applied to propose, discuss, and test assumptions about the characteristics and capabilities of end users of eHealth systems and applications.
- During participatory design, the selection of participants (users) can take into account the level of eHealth literacy possessed by a broad range of users of the proposed system.
- The potential range of users, defined in terms of eHealth literacy capability, is considered explicitly when modeling user requirements and/or more specifically envisioning classes of end users of the system or application.

Although it is acknowledged that eHealth literacy *is not* the only important dimension or factor that needs to be considered in the development of effective consumer-focused health IT systems and applications, we argue that it is foundational and a critical success factor and that *not* considering end-user eHealth literacy levels may increase safety risks and/or risks of poor outcomes. Other factors in design are consideration of user-specific information architectures and navigational issues, with a growing body of literature addressing the options regarding choices for how to best display information to eHealth consumers and development of guidelines for creating eHealth applications; however, this is out of scope of this paper. The next section illustrates how the framework can be utilized during requirements gathering for the development of consumer-focused eHealth systems and/or applications.

Figure 1. User-task-context matrix for health information technology systems design (adapted from Kushniruk and Turner [23]).

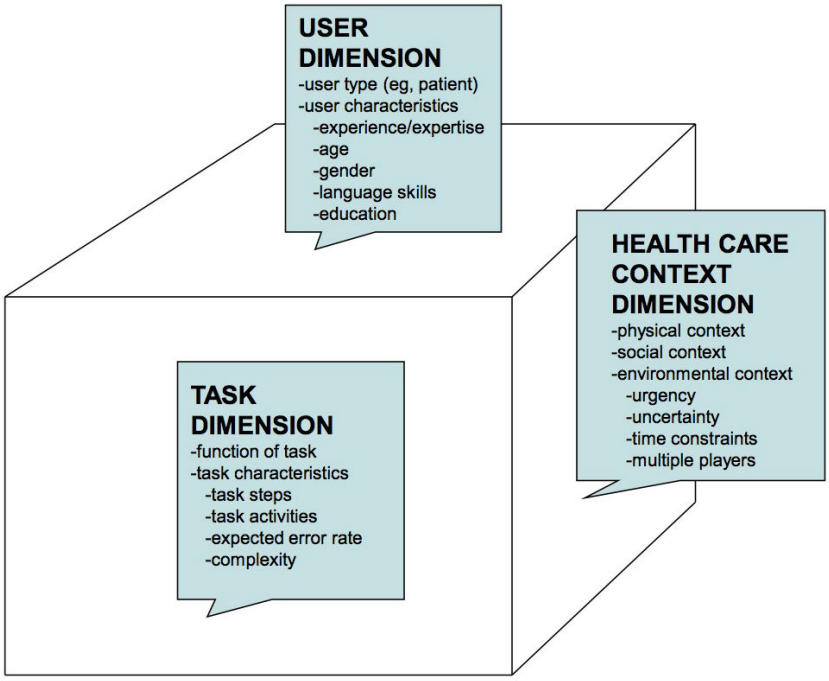
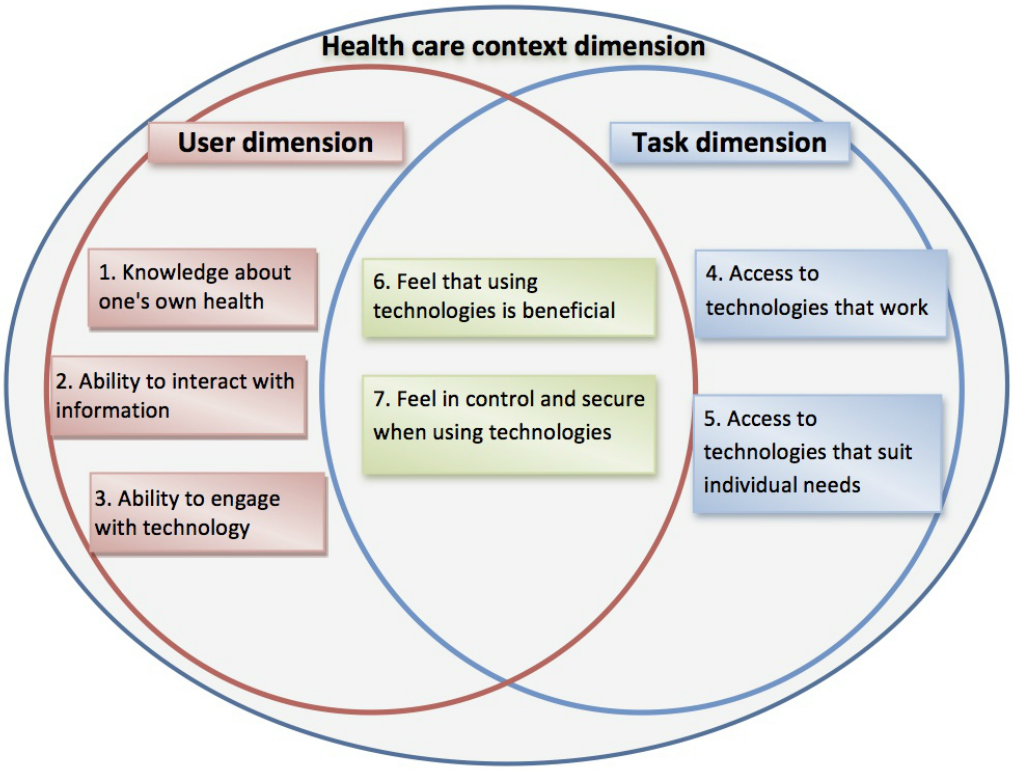


Figure 2. Framework for Design: Expanded User-task-context matrix incorporating eHealth literacy.



Results

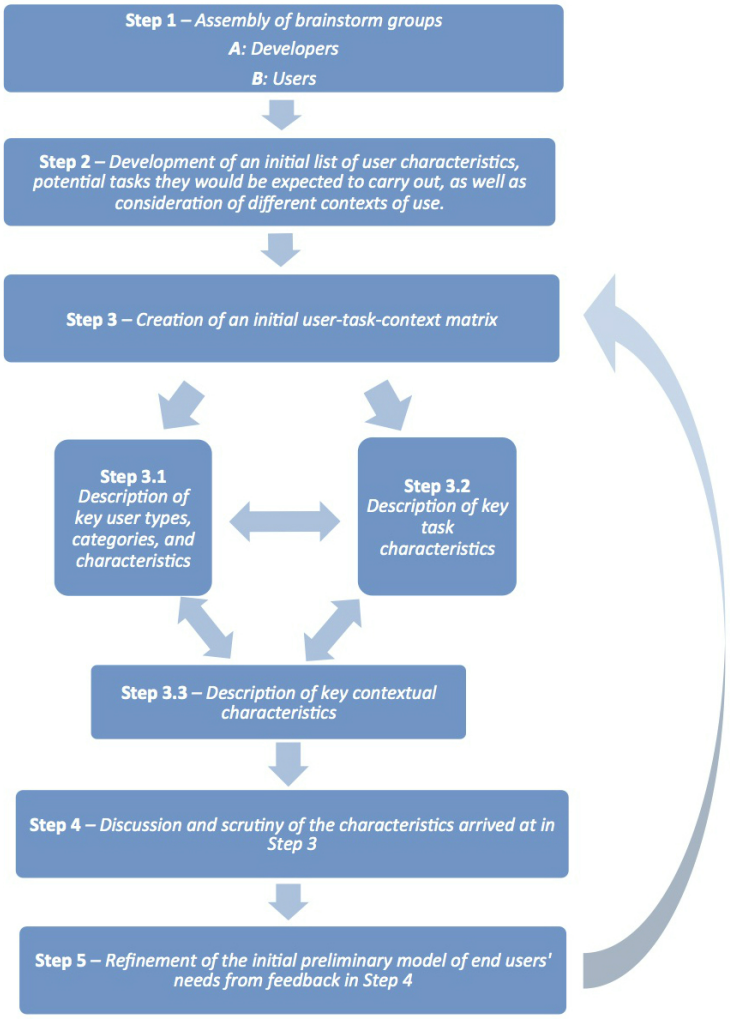
An Iterative Five-Step Framework for Understanding Users' Needs

Overview

The aforementioned description illustrates that it is possible to integrate the new eHealth literacy concept (and its 7 domains) with an expanded user-task-context matrix to describe and qualify user requirements and understanding. This section presents the resulting framework that has been produced and structured as a five-step process aimed at supporting eHealth system designers to improve their understanding and more accurately address end-users' needs, capabilities, and contexts of use to improve the effectiveness and broader applicability of consumer-focused eHealth systems and applications.

The following description presents the framework and each of the five steps within it. It highlights that within the framework, generic requirements-gathering approach can be used to enhance understanding, modeling, and reasoning about end-user profiles. Providing designers with personas and/or vignettes can enable them to visualize and understand the particular needs of the span of users. Similarly, specification of requirements can fit into current innovation models such as participatory design [28], rapid contextual design [29], or the biodesign innovation process [30]. The approach can also be recommended when known technologies are planned to be applied in new settings to ensure that the technology can meet users' capabilities and expectations. The iterative five-step framework includes and expands on the basic steps described by Hackos and Redish [24] for developing a user-task matrix: specifically starting with brainstorming a list of users, creating an initial matrix, and then testing assumptions about users. In the following description, these steps are elaborated to include the consideration of a third dimension—that of context, and to also explicitly highlight consideration of eHealth literacy within the framework when modeling users of health technologies. The steps involved are illustrated in Figure 3 and are described as follows:

Figure 3. Information technology requirements gathering flowchart.



Step 1: Assembly of a Brainstorm Group

Overview

This step consists of two processes. The first involves constituting a group of professionals capable of leading the whole process of requirements gathering. The size of this group should be approximately 8 persons as this will be easier to create a team with trust among the participants. The second involves recruitment of a reference panel.

Developers

An initial group with competences within the area for which the requirements are built is assembled. This group should consist of people trained in health informatics, computer science, health care, and behavioral sciences. This group will lead the process including recruitment of a reference panel consisting of end users.

User Panel

Recruitment of a reference panel. This panel is most effective when the members are typical end users (laypeople), rather than informed and engaged patients, such as those who may already be active members of patient organizations. The panel will be used in Steps 3.3 and 5 and ideally would consist of 50-150 persons with very wide eHealth literacy levels. However, it should be noted that practical constraints may limit the number of available participants to a smaller number.

Step 2: Development of an Initial List of User Characteristics, Potential Tasks They Would Be Expected to Carry Out, as well as Consideration of Different Contexts of Use

For each user, the profile, role, and rights are described. A role can be doctor, nurse, patient, relative, citizen, for example.

In a second iteration, a similar approach can be used with respect to the rights in the system (eg, administrator, superuser, user, supporter). However, in most systems and applications in health care, the professional identity is more important than the rights because consumers, patients, and citizens will expect more rights, and the ability to grant permission to other actors, such as their health professionals and relatives.

In a third iteration, these initial profiles should be aligned with the tasks to be solved and the context (eg, how should the individual profiles act together as collaborators or providers and receivers of services). A key aspect of Step 2 is the preliminary identification of the different categories of end users of the system or application and development of user profiles for each of these categories.

Step 3: Creation of an Initial User-Task-Context Matrix

Overview

Preliminary ideas from Step 2 are formalized across the user-task-context matrix formalism (as shown in Figure 1). As there are three main dimensions, this step involves three steps (Steps 3.1, 3.2, and 3.3).

Steps 3.1 and 3.2 provide a description of how the associated domains of eHealth literacy are used to qualify the understanding of needs and how a requirement specification is produced, respectively. In Step 3.3, an evaluation of the output from Steps 3.1 and 3.2 is considered in relation to the overarching framework.

It is important to note that the seven eHealth literacy domains together describe the requirements for a successful solution to be offered to users with a high digital confidence (ie, knowledge, skills, and motivation in the context of health). By contrast, the solution will also be beneficial for a wider audience if, for each domain, it is considered regarding how the system can adapt to and handle users with lower knowledge levels, skills, and motivation, by taking design, functionality, and assistive technologies or support tools into consideration.

For this purpose, the seven eHealth domains are qualified for use in the three areas by designing personas with high and low eHealth literacy, respectively, within each of the domains. These 14 descriptors are used in the relevant sections to ensure that the developer team can envisage the typical user span, to ensure the system or application will meet the full range of users.

Step 3.1: Description of Key User Types, Categories, and Characteristics

Capabilities

Knowledge About One's Own Health

Need to consider how users can be supported if they do not know the terms or functions related to their health, how IT systems ensure that the users do not misinterpret information, and how users can be supported to learn more about the specific health area that the system or program intends to cover.

Do links to resources about health and the health care system clearly explain key features in relation to the tasks? What kinds of resources are needed to involve users with low abilities in this area?

Ability to Interact With Information

Need to consider how any system addresses the needs of people with low ability to read or calculate or orientate themselves within complex designs. How can the solution assist dyslectics by voice or video and how can figures and calculations be represented in a way that people with dyscalculia can handle?

Ability to Engage With Technology

The key questions in this step include how the user interface can address users with different knowledge of how to access systems and interact with them.

The log-in procedures should be simple, the system design should be directed toward integration with other systems, which the typical users are expected to use. The interface should be familiar to other systems and be designed according to international recommendations, standards, and the look of the most used platforms such as Windows and Mac operating systems. The tasks or system should also be easily accessible using various devices and technologies.

Step 3.2: Description of Key Task Characteristics

Overview

Describe the tasks that different categories of users are expected to carry out through interaction with the system or application

being developed. For example, if a task requires input of numeric data, this may impose additional requirement demands on the user being low in domain 2, whereas a task that requires inputs from quick response (QR) codes will not impose numerical challenges. Another example is input or access to personal health data, and how this might influence whether or not the user feels safe and in control with the system. A relevant question is, "Can the task be designed to meet the users' capabilities or should alternative tasks be created?"

Access to Technologies

Access to Technologies That Work

How does the system or application document its stability and uptime. After how long a period do the users experience that the system is not accessible? What is the critical uptime to work sufficiently and what is the perspective of the various user groups on access availability?

Access to Technologies That Suit Individual Needs

The key characteristics are the usability of the system, adaptability to users with particular needs, and opportunities to collaborate with others if needed.

The interface should be user friendly and designed to be configurable and adaptable to how it is used. Individuals with particular needs due to disabilities should be addressed to avoid inequity in access to the system.

Assistance as built-in support and/or possibilities of getting help from friends, relatives, peers, or health professionals should be considered in the design.

Personas with various disabilities should be created with a focus on the kinds of users who are intended to use the system and which types of disabilities might be expected. Which kinds of individual needs will the typical users have with respect to accessibility, mobility, and variability of interfaces? To what extent will the system need to tailor to individuals with a wide range of individuals' needs or disabilities?

Next, the tasks are described from the end-user experience perspective.

Experience Using Technologies

Feel That Using Technologies Is Beneficial

How might the system or application be experienced as beneficial compared with known solutions or other digital products? How can experiences during the use of the solution induce motivation in users with low motivation; for example, through gaming or experienced benefits or victories? How can users with low belief of a new product or accustomed to current solutions be met and taken care of by the solution?

Feel in Control and Secure When Using Technologies

How can the users know who has access? How can the interface promote confidence in the system's security to ensure that people accessing personal data are only those granted permission by the user or, if not, it is clear to the user how and when others access their data? The system should aspire to meet the needs of users ranging from those with a high degree of confidence to those who are concerned about this topic. It should be considered how data management can be transparent and how data are used for users with low confidence or a high degree of mistrust.

Step 3.3: Description of Key Contextual Characteristics

As described earlier, the issue of context is critical for developing health care applications that are likely to be used effectively and adopted by end users.

In this step, contexts of use should be identified based on observations in real-life environments and interviews with users and compared with the outputs from Steps 3.1 and 3.2, and should be qualified with respect to identified contextual issues.

The field for observations as well as the users should be identified in Step 1 (subcategory Users) and the users should be recruited in a way that ensures a broad representation of eHealth literacies with variations across domains. The variances can be identified based on semistructured interviews containing questions relating to the descriptors of the domains. It is expected that questionnaires to be applied will be available by 2016. The number of users needed for a panel will vary depending on resources for the development, but the number should be representative for at least one being high and one being low in each of the domains. This can be accomplished by selecting 14 individuals representing the 14 personas described earlier with respect to either high or low level within each of the seven ehealth domains. It will need 128 (2^7) if a full matrix of variations should be sampled. However, it will probably require testing of a considerably larger population. Therefore, it is recommended to focus on profiles that appear to be the most common within a sample of users. This sample is ideally recommended to be within 50-300 depending on resources (although practical constraints may limit the number to the lower recommended number). For the observed typical combinations of user and task, narratives or fleshed-out use cases should be developed to better define user capabilities in terms of expected tasks. These use cases become an expression of user requirements for systems being developed. They can also be used upon completion of working or partially working prototype versions of the system for setting user- and usability-testing scenarios.

In developing applications and systems for health care professionals, this has typically involved understanding of whether there are contextual issues that would render a system unusable under certain conditions (eg, are there environments or settings where a system would not work?), while other conditions might be seen as being conducive for use of the system. For example, a speech recognition interface to a health information system may work well in a private office, but not at all in the context of a noisy clinical environment. With regard to eHealth applications and systems targeted at patients, laypeople, and the general population, the contextual issue becomes magnified, with an even greater range of settings, situations, and contexts in which a system might be used. To fully understand requirements for a new system or application, an upfront understanding of such contextual facilitators or inhibitors of a system or application is needed.

Step 4: Discussion and Scrutiny of the Characteristics Arrived at in Step 3

In this step, the model of users and their interactions with the system in terms of tasks and contextual factors is discussed and scrutinized by the development team. As this will represent only an initial or preliminary model of end users, plans for studying the assumptions and testing them empirically using the "user panel" from Step 1 have to be made. In this step, each of the seven eHealth literacy domains should be taken into account by focusing on how the system meets and fulfills the needs of personas, described in Steps 3.1-3.3. The most important issue here is that the developers undertake the iterations from Steps 3.1-3.3 and ensure that the choices made here remain consistent with the planned requirements for the solution and its user interface, functions, and assistive technologies.

At this stage of development, the designers and developers will test prototypes presenting the content and various ways to interact with it. Here it is important to involve the wide range of identified personas. This may involve discussions with end users, application of contextual enquiry, and usability testing.

Step 5: Refinement of the Initial Preliminary Model of End-Users' Needs from Feedback in Step 4

The model of the end-users' need developed and tested should be considered in light of feedback from Step 4 by testing the model's main assumptions. Although developers may hesitate to spend the time and effort needed to refine the initial model represented by the user-task-context matrix, leaving this step out will increase the likelihood that the system or application ultimately developed will not meet the needs of the end users. The refined model can then be used to drive further requirements gathering and streamline the development of system use cases during the design phase and later on during the implementation and testing phases. The model can also be used to create test scenarios.

A Case Story

As an example of how detailed knowledge about individuals, including eHealth literacy, and the specific context can populate the framework, we briefly present a patient case and describe how considerations related to the seven eHealth literacy domains can feed into Steps 3.1 and 3.2 of the framework. The case was constructed from characteristics identified from a number of real patients in the Epital project [31-33] and author LK's clinical experience.

The Epital example is chosen as this project consists of a redesign of health services using a whole system approach, service transformation, user centeredness, and especially digital support for both the enrolled people with chronic obstructive pulmonary disease (COPD) and the health care providers.

> *Angela is a 72-year-old woman with COPD that was diagnosed 5 years ago. Although she takes her medications regularly, she does have exacerbations about 4 times a year and is in need of hospitalization or referral to an emergency room 2 times a year. Because of the condition, she has anxiety, is slightly cognitively impaired, especially when exacerbations are emerging, and it is at this particular time she needs to use her Epital navigator to connect to her digital service and trust that it works instantly. Owing to treatment of her COPD with beta-2-adrenergic stimulators, she suffers from tremor of her hands, which is exaggerated when her condition worsens, partly due to an increased intake of medicine and partly due to an increase in anxiety. She has an average score in self-reported health literacy, only few errors in a test of health literacy primarily related to calculations of medicine doses, whereas her understanding of measurements related to her condition is found to be insufficient and makes it difficult for her to interpret and act on data outside the normal range. Although she scores on average in most tests, the confidence and speed of performance are not good.*

Table 1 describes how this profile can be utilized when designing a home monitoring solution for patients with COPD.

Table 1. Using the seven-domain eHealth concept in Health information technology design.

eHealth literacy domains	Description
Knowledge about one's own health (Domain 1)	The patient has sufficient basic information but her understanding of more complex data makes it difficult to interpret her own measurements. Here it should be considered to build in a function that assists interpretation.
Ability to interact with information (Domain 2)	Angela is able to read but has some problems with calculation, which may be taken into consideration here.
Ability to engage with technology (Domain 3)	She scores average in confidence and skills with computers, and thus, the solution should either include training for the particular system or assisting functions should be included, which will address the areas in which the patient is insecure and has a low digital competence.
Access to technologies that work (Domain 4)	Her condition could, in a short time, be life threatening and it is therefore important to allow for feedback about connectivity and a way to secure backups to avoid an increase of her anxiety, which is often related to a lack of confidence in health information technology services.
Access to technologies that suit individual needs (Domain 5)	She has tremor of her hands and needs a very simple button or a technology like pointers to be sure that she presses the right areas. She will also need a solution that provides interaction with real persons when her exacerbations are severe because she will not be able to interact with the technology without voice or video support.
Feel that using technologies is beneficial (Domain 6)	Angela has been introduced to a home monitoring device in the form of a tablet computer. It should be considered how a simple user interface can be developed because she has felt insecure when she performs tasks indicating uncertainty in relation to the technology and her own condition. This may be influenced by her medicine and possible anxiety state.
Feel in control and secure when using technologies (Domain 7)	She has scored as being confident about the Internet and the use of computers, and therefore, this domain may not be of concern.

Discussion

The proposed expanded framework has been designed to provide a user-friendly framework for design. It will directly assist health systems policy development and decision making because it can be deployed to ensure that the design process takes the users' capabilities, and the circumstances in which resulting products are to be used, into consideration. This will maximize the probability of generating usable, safe, and highly implementable IT applications and systems, and generate the intended positive health outcomes. From previous work, it has been found that fitting health information system design to users' understanding and capabilities leads to a thoughtful design and more usable and useful systems and that ease of use and usefulness are also associated with not just more effective systems, but also with safer systems [2]. It should be noted that the approach described in this paper gives an overall framework for considering users of health information technologies to highlight consideration of eHealth literacy explicitly. As such, it will complement specific data-collection methods (eg, the use of the think-aloud method or simulations) used in obtaining detailed information about users, their needs, and their capabilities [26]. In addition, the approach can be applied not only for design, but also for documentation of eHealth application requirements that can be consulted once an application is developed to help target end-user support (eg, by assessing whether individuals of differing user classes can do the tasks the eHealth application was designed to support once in widespread use). It should also be noted that the approach could be applied in the development of eHealth applications targeted for individuals as well as for groups. For example, in development of regional or national personal health record and personal health portals, consideration is needed of what the major classes of users of the application are, their eHealth literacy, and the type of tasks they would find useful (including collaborative and socially oriented tasks) to perform using the application. As there is a move toward greater use of social media this need will only increase.

The proposed new concept of eHealth literacy [34] has similarities to Norman and Skinner's model with six core literacies [10] within the areas of tradition alliteracy, health literacy, computer literacy, information literacy, similar to the domains "1. Knowledge about one's own health," "2. Ability to interact with information," and "3. Ability to engage with technology." The media and science literacy are not represented as strongly as in our new concept. By contrast, based on the statements in the development process, we have got a unique insight into how the users (ie, patients, health professionals, and technology developers) are thinking. These insights can be used as a framework to describe where developers need to be aware of in the design processes as it is presented in this paper.

Thus, we are able to present an expanded model of the user-task-context matrix, which qualifies needs to be addressed not only with respect to the users, but also how they expect tasks and systems to be designed and also how the intersection between the users' competencies and the service design should be addressed to motivate the users, having them feel safely in control, and being able to interact. In this way, our concept with its underlying statements contributes to a new understanding of eHealth literacy. The concept also includes subjects previously identified to be essential to ensure usage of health technology [35,36], that is, trust that is embedded in the domains "4. Access to technologies that work" and "7. Feel in control and secure when using technologies and interpersonal relationships," which is embedded in the domain "5. Access to technologies that suit individual needs" where statements from the development process cover aspects of being able to share data with relatives, support relatives, or receive support from others. This aspect is also shortly described in our recommendations in Step 3.2. We find that our seven domain eHealth literacy concepts covers both the important domains

of the definition and the model by Norman and Skinner. At the same time, it also covers other dimensions of importance for understanding the interaction between users and the digital services offered by health care providers in the future.

This new eHealth literacy concept has turned out to have similar focus on several areas of the user-task-context matrix suggested by Kushniruk and Turner [23]. Because this matrix lacked dimensions for user characteristics related to eHealth literacy, we have integrated the concept of eHealth literacy and the matrix into an expanded framework, which can be used to produce requirements for designers. It is anticipated that the framework presented directly supports the more accurate identification of differences in users, and their contexts of use, and how these factors interact with usability and the risk of unintended consequences from health IT systems. The new concept of eHealth literacy and the expanded framework presented suggest that there are ways for opening up new conversations about innovative ways of thinking, designing, and empowering health consumers in their use of health information and health information systems. The expanded framework contains personas and vignettes/narratives, which illustrate the needs of users with both high and low competences.

To maximize the impact of the framework, we have specified its application within a structured process that involves a multidisciplinary small group to facilitate and lead the process and to create a user panel of laypeople to ensure connection to the real world. In this way, the presented framework can expand how users are involved in the initial phases of other innovation models such as the participatory design [28] or rapid contextual design [29]. The proposed process systematically qualifies them through a grounded understanding of the mandatory areas that must be understood and anticipated to maximize usability from both an end-user and a system approach. The framework will also be of great importance when the Stanford Biodesign process is used in development of eHealth solutions. The first step of the framework will be skipped at this point in the Biodesign process because cross-disciplinary teams have already been established. Steps 2-5 can be embedded in the invention phase when prototyping begins to ensure that the involved stakeholders and patient segments are understood and addressed with respect to eHealth literacy, and that specific criteria are included in the need statements to ensure a safe and usable product.

Although there is extensive evidence that health literacy is associated with health outcomes [37], and that eHealth interventions may improve health outcomes [38], there is limited evidence connecting eHealth literacy and the use of eHealth interventions. One reason for this is that in many cases projects are not designed on a large scale and may not disseminate knowledge based on evidence [39].

Before the full benefit of the framework can be obtained, methods to estimate individuals' eHealth literacy are required. At present, we are developing such tools, which are expected to be available by the end of 2015. At this point, it is recommended to use semistructured interviews based on the descriptors in Textbox 1 to classify users according to their eHealth literacy.

For developers, it is expected that the proposed expanded framework for design will ensure that future health IT products involving innovation, design, and maturation phases will be designed to accommodate the needs of a variety of users from a system-design approach involving eHealth literacy. The intention is to ensure a robust and safe system where the developers involved will also understand the situations and contexts in which the system will be used. Although the entire expanded framework has not yet been implemented in full, we invite designers to use it and to create research projects where its effects can be documented.

Developing effective eHealth technologies requires an understanding of the needs of end users from multiple perspectives. In this paper, we have proposed and detailed a framework for modeling users' needs for eHealth that merges prior work in development of a user-task-context matrix with the emerging area of eHealth literacy. This framework is intended to be used to guide design of eHealth technologies and to make requirements explicitly related to eHealth literacy, enabling a generation of well-targeted, fit-for-purpose, equitable, and effective products and systems.

Acknowledgments

The authors thank Ms Lena Sundby Jensen and Ms Melanie Hawkins for their assistance in editing this paper. Parts of the project took place when LK and ON were funded by the Danish National Strategic Research Council. RO was funded in part by a National Health and Medical Research Council of Australia Senior Research Fellowship.

Conflicts of Interest

None declared.

References

1. Kujala S. User involvement: A review of the benefits and challenges. Behav Inf Technol 2003 Jan;22(1):1-16. [doi: 10.1080/01449290301782]
2. Kushniruk AW, Bates DW, Bainbridge M, Househ MS, Borycki EM. National efforts to improve health information system safety in Canada, the United States of America and England. Int J Med Inform 2013 May;82(5):e149-e160. [doi: 10.1016/j.ijmedinf.2012.12.006] [Medline: 23313431]

3. Schramm W. How communication works. In: The Process and Effects of Mass Communication. Urbana, IL: University of Illinois Press; 1954:3-26.
4. 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):259-269. [doi: 10.1093/heapro/15.3.259]
5. Norman CD, Skinner HA. eHealth literacy: Essential skills for consumer health in a networked world. J Med Internet Res 2006;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
6. International Organization for Standardization. Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs). ISO 9241-11:1998. Geneva, Switzerland: International Organization for Standardization; 1998. Part 11: Guidance on Usability URL: http://www.iso.org/iso/catalogue_detail.htm?csnumber=16883 [accessed 2015-05-13] [WebCite Cache ID 6YUerjeMM]
7. Cho AH, Arar NH, Edelman DE, Hartwell PH, Oddone EZ, Yancy WS. Do diabetic veterans use the internet? Self-reported usage, skills, and interest in using My HealtheVet Web portal. Telemed J E Health 2010 Jun;16(5):595-602. [doi: 10.1089/tmj.2009.0164] [Medline: 20575727]
8. Hargittai E. Survey measures of web-oriented digital literacy. Soc Sci Comput Rev 2005 Aug 01;23(3):371-379. [doi: 10.1177/0894439305275911]
9. van Deursen AJ, van Dijk JA. Internet skills performance tests: Are people ready for eHealth? J Med Internet Res 2011;13(2):e35. [doi: 10.2196/jmir.1581] [Medline: 21531690]
10. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006;8(4):e27. [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
11. Norman C. eHealth literacy 2.0: Problems and opportunities with an evolving concept. J Med Internet Res 2011;13(4):e125. [doi: 10.2196/jmir.2035] [Medline: 22193243]
12. Ellis L, Showell C, Turner P. Social media and patient self-management: Not all sites are created equal. Stud Health Technol Inform 2013;183:291-295. [Medline: 23388301]
13. 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;15(2):e30. [doi: 10.2196/jmir.2237] [Medline: 23406655]
14. Hanson CL, Cannon B, Burton S, Giraud-Carrier C. An exploration of social circles and prescription drug abuse through Twitter. J Med Internet Res 2013;15(9):e189. [doi: 10.2196/jmir.2741] [Medline: 24014109]
15. Buchbinder R, Batterham R, Elsworth G, Dionne CE, Irvin E, Osborne RH. A validity-driven approach to the understanding of the personal and societal burden of low back pain: Development of a conceptual and measurement model. Arthritis Res Ther 2011;13(5):R152 [FREE Full text] [doi: 10.1186/ar3468] [Medline: 21933393]
16. Osborne RH, Batterham RW, Elsworth GR, Hawkins M, Buchbinder R. The grounded psychometric development and initial validation of the Health Literacy Questionnaire (HLQ). BMC Public Health 2013;13:658 [FREE Full text] [doi: 10.1186/1471-2458-13-658] [Medline: 23855504]
17. Osborne RH, Elsworth GR, Whitfield K. The Health Education Impact Questionnaire (heiQ): An outcomes and evaluation measure for patient education and self-management interventions for people with chronic conditions. Patient Educ Couns 2007 May;66(2):192-201. [doi: 10.1016/j.pec.2006.12.002] [Medline: 17320338]
18. Osborne RH, Norquist JM, Elsworth GR, Busija L, Mehta V, Herring T, et al. Development and validation of the Influenza Intensity and Impact Questionnaire (FluiiQ™). Value Health 2011 Aug;14(5):687-699. [doi: 10.1016/j.jval.2010.12.005] [Medline: 21839407]
19. Busija L, Buchbinder R, Osborne RH. A grounded patient-centered approach generated the personal and societal burden of osteoarthritis model. J Clin Epidemiol 2013 Sep;66(9):994-1005. [doi: 10.1016/j.jclinepi.2013.03.012] [Medline: 23774110]
20. Jordan JE, Buchbinder R, Briggs AM, Elsworth GR, Busija L, Batterham R, et al. The Health Literacy Management Scale (HeLMS): A measure of an individual's capacity to seek, understand and use health information within the healthcare setting. Patient Educ Couns 2013 May;91(2):228-235. [doi: 10.1016/j.pec.2013.01.013] [Medline: 23419326]
21. Trochim WM, Cook JA, Setze RJ. Using concept mapping to develop a conceptual framework of staff's views of a supported employment program for individuals with severe mental illness. J Consult Clin Psychol 1994 Aug;62(4):766-775. [Medline: 7962880]
22. Trochim MK. Concept Mapping for Planning and Evaluation. Thousand Oaks, CA: Sage Publications; 2007.
23. Kushniruk A, Turner P. A framework for user involvement and context in the design and development of safe e-Health systems. Stud Health Technol Inform 2012;180:353-357. [Medline: 22874211]
24. Hackos JT, Redish JC. User and Task Analysis for Interface Design. New York: Wiley; 1998.
25. Rasmussen R, Kushniruk A. Digital video analysis of health professionals' interactions with an electronic whiteboard: A longitudinal, naturalistic study of changes to user interactions. J Biomed Inform 2013 Dec;46(6):1068-1079 [FREE Full text] [doi: 10.1016/j.jbi.2013.08.002] [Medline: 23954312]
26. Li AC, Kannry JL, Kushniruk A, Chrimes D, McGinn TG, Edonyabo D, et al. Integrating usability testing and think-aloud protocol analysis with "near-live" clinical simulations in evaluating clinical decision support. Int J Med Inform 2012 Nov;81(11):761-772. [doi: 10.1016/j.ijmedinf.2012.02.009] [Medline: 22456088]

27. Chan CV, Kaufman DR. A framework for characterizing eHealth literacy demands and barriers. J Med Internet Res 2011;13(4):e94. [doi: 10.2196/jmir.1750] [Medline: 22094891]
28. Bødker K. Participatory IT Design: Designing for Business and Workplace Realities. Cambridge, MA: MIT Press; 2004.
29. Holtzblatt K. Rapid Contextual Design: A How-To Guide to Key Techniques for User-Centered Design. San Francisco, CA: Elsevier/Morgan Kaufmann; 2005.
30. Zenios SA, Makower J, Yock PG. Biodesign: The Process of Innovating Medical Technologies. Cambridge, UK: Cambridge University Press; 2010.
31. Vang A, Knudsen A. Fokus på e-health Literacy ved Hjemmemonitorering af KOL-Borgere Internet Thesis from the BSc Health Informatics Program. University of Copenhagen: Copenhagen, Denmark; 2014. URL: http://dias.kb.dk/downloads/dias:128?locale=da [accessed 2015-05-13] [WebCite Cache ID 6YUfWZ4P1]
32. Phanareth K, Christensen AS, Vingtoft S, Kayser L. Int J Integr Care. 2013. Epital Health: A disruptive and research based approach to service transformation of health care systems URL: http://www.ijic.org/index.php/ijic/article/view/URN%3ANBN%3ANL%3AUI%3A10-1-115733 [accessed 2015-05-13] [WebCite Cache ID 6YUfbgKvA]
33. Hesseldal L, Kayser L. Healthcare innovation - The Epital: A living lab in the intersection between the informal and formal structures. Qual Sociol Rev 2016;12 (forthcoming).
34. Kayser L, Osborne RH, Nørgaard O, Turner P. E-Sundhedskompetence - et Redskab til at Skabe Bro Mellem Borgere, Patienter og Vores Digitale Sundhedstilbud Internet. 2014 Oct 2. URL: http://2014.e-sundhedsobservatoriet.dk/sites/2014.e-sundhedsobservatoriet.dk/files/slides/Lars%20Kayser%2C%20C1.pdf [accessed 2015-05-13] [WebCite Cache ID 6YUfoBl8p]
35. Kontos EZ, Bennett GG, Viswanath K. Barriers and facilitators to home computer and internet use among urban novice computer users of low socioeconomic position. J Med Internet Res 2007;9(4):e31. [doi: 10.2196/jmir.9.4.e31] [Medline: 17951215]
36. Montague E, Asan O. Trust in technology-mediated collaborative health encounters: Constructing trust in passive user interactions with technologies. Ergonomics 2012;55(7):752-761 [FREE Full text] [doi: 10.1080/00140139.2012.663002] [Medline: 22506847]
37. 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]
38. Steventon A, Bardsley M, Billings J, Dixon J, Doll H, Hirani S, Whole System Demonstrator Evaluation Team. Effect of telehealth on use of secondary care and mortality: Findings from the Whole System Demonstrator cluster randomised trial. BMJ 2012;344:e3874 [FREE Full text] [doi: 10.1136/bmj.e3874]
39. Jakobsen NK, Jensen LS, Kayser L. Collaborative efforts are needed to ensure proper knowledge dissemination of telemedicine projects. Dan Med J 2014 Sep;61(9):A4896. [Medline: 25186538]

Abbreviations

COPD: chronic obstructive pulmonary disease
eHEALS: eHealth Literacy Scale
IT: information technology
QR: quick response

Edited by G Eysenbach; submitted 16.07.14; peer-reviewed by M Stellefson, P Boisrond, T Veinot; comments to author 19.10.14; revised version received 15.01.15; accepted 26.04.15; published 20.05.15.

Please cite as:
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 Human Factors 2015;2(1):e9
URL: http://humanfactors.jmir.org/2015/1/e9/
doi:10.2196/humanfactors.3696
PMID:27025228

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org